

BA

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification G: <b>C12Q 1/68</b>		<b>A2</b>	(11) International Publication Number: <b>WO 99/14375</b>
			(43) International Publication Date: <b>25 March 1999 (23.03.99)</b>
(21) International Application Number: <b>PCT/US98/19576</b>		(74) Agent: MULLER, M., Michèle; Arnold, White & Durkee, P.O. Box 4423, Houston, TX 77210 (US).	
(22) International Filing Date: <b>18 September 1998 (18.09.98)</b>			
(30) Priority Data: <b>60-059,415 19 September 1997 (19.09.97) US</b>		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GR, GM, GU, HK, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Parisian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application <b>US 60-059,415 (CIP) Filed on 19 September 1997 (19.09.97)</b>			
(71) Applicant (for all designated States except US): GENETRACK SYSTEMS, INC. [US/US]; 1401 Harbor Bay Parkway, Alameda, CA 94502 (US).		Published <i>Without international search report and to be republished upon receipt of this report.</i>	
(72) Inventors; and (73) Inventor/Applicants (for US only): RUTHER, John, M. [US/US]; Apartment D, 858 Colman Avenue, Menlo Park, CA 94025 (US); LI, Jia [CN/US]; 32801 Regents Boulevard, Union City, CA 94587 (US); MONROTE, Joseph [US/US]; 50 Alamo Avenue, Berkeley, CA 94708 (US); RECKNER, Christopher, A. [US/US]; 106 Clover Lane, Menlo Park, CA 94025 (US).			

(54) Title: DNA TYPING BY MASS SPECTROMETRY WITH POLYMORPHIC DNA REPEAT MARKERS

(57) Abstract

The present invention is related to the fields of genetic mapping and genetic identity detection, including forensic identification and paternity testing. This invention is more specifically directed to the use of mass spectrometry to detect length variation in DNA nucleotide sequence repeats (including variants of common alleles), such as microsatellites and short tandem repeats, and to DNA sequence provisors as primers for the analysis of DNA tandem nucleotide repeat polymorphisms at specific loci on specific chromosomes.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AI.	Algeria	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Algeria	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GI	Gibraltar	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	ML	Mali	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	MM	Myanmar	TU	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Morocco	UA	Ukraine
BN	Brunei	IL	Israel	MY	Malaysia	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KR	Republic of Korea	NO	Norway	YU	Yugoslavia
CI	Cote d'Ivoire	KG	Kyrgyzstan	NZ	New Zealand	ZM	Zimbabwe
CM	Cameroon	KP	Democratic People's Republic of Korea	PL	Poland		
CN	China	KZ	Kazakhstan	PT	Portugal		
CU	Cuba	LA	Laos	RO	Romania		
CZ	Czech Republic	LC	St. Lucia	RU	Russian Federation		
DK	Denmark	LJ	Liechtenstein	SD	Sudan		
DM	Dominica	LN	St. Luce	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

## **DESCRIPTION**

### **DNA TYPING BY MASS SPECTROMETRY WITH POLYMORPHIC DNA REPEAT MARKERS**

#### **BACKGROUND OF THE INVENTION**

##### **5 A. Field of the Invention**

The present invention is generally directed to the field of genetic identity detection including forensic identification and paternity testing as well as genetic mapping. The present invention is more specifically directed to the use of mass spectrometry to detect length variations in DNA nucleotide sequence repeats, often referred to as short tandem repeats ("STR"), microsatellite repeats or simple sequence repeats ("SSR"). The invention is also directed to DNA sequences provided for the analysis of STR polymorphisms at specific loci on specific chromosomes.

10

##### **B. Description of Related Art**

Polymorphic DNA tandem repeat loci are useful DNA markers for paternity testing, human identification, and genetic mapping. Higher organisms, including plants, animals and humans, contain segments of DNA sequence with variable sequence repeats. Commonly sized repeats include dinucleotides, trinucleotides, tetranucleotides and larger. The number of repeats occurring at a particular genetic locus vary depending on the locus and the individual from a few to hundreds. The sequence and base composition of repeats can vary significantly, not even remaining constant within a particular nucleotide repeat locus. DNA nucleotide repeats are known by several different names including microsatellite repeats, simple sequence repeats, short tandem repeats and variable nucleotide tandem repeats. As used herein, the term "DNA tandem nucleotide repeat" ("DTNR") refers to all types of tandem repeat sequences.

15

20

25

Thousands of DTNR loci have been identified in the human genome and have been predicted to occur as frequently as once every 15 kb. Population studies have been undertaken on dozens of these STR markers as well as extensive validation studies in forensic laboratories. Specific primer sequences located in the regions flanking the DNA tandem repeat region have been used to amplify alleles from DTNR loci *via* the polymerase chain reaction ("PCR™"). Thus, the PCR™ products include the polymorphic repeat regions, which vary in length depending on the number of repeats or partial repeats, and the flanking regions, which are typically of constant length and sequence between samples.

The number of repeats present for a particular individual at a particular locus is described as the allelic value for the locus. Because most chromosomes are present in pairs, PCR™ amplifications of a single locus commonly yields two different sized PCR™ products representing two different repeat numbers or allele values. The range of possible repeat numbers for a given locus, determined through experimental sampling of the population, is defined as the allelic range, and may vary for each locus, e.g., 7 to 15 alleles. The allele PCR™ product size range (allelic size range) for a given locus is defined by the placement of the two PCR™ primers relative to the repeat region and the allelic range. The sequences in regions flanking each locus must be fairly conserved in order for the primers to anneal effectively and initiate PCR™ amplification. For purposes of genetic analysis di-, tri-, and tetranucleotide repeats in the range of 5 to 50 are typically utilized in screens.

Many different primers have been designed for various DTNR loci and reported in the literature. These primers anneal to DNA sequences outside the DNA tandem repeat region to produce PCR™ products usually in the size range of 100-800 bp. These primers were designed with polyacrylamide gel electrophoretic separation in mind, because DNA separations have traditionally been performed by slab gel or capillary electrophoresis. However, with a mass spectrometry approach to DTNR typing and analysis, examining smaller DNA oligomers is advantageous because the sensitivity of detection and mass resolution are superior with smaller DNA oligomers.

The advantages of using mass spectrometry for characterizing DTNRs include a dramatic increase in both the speed of analysis (a few seconds per sample) and the accuracy of direct mass measurements. In contrast, electrophoretic methods require significantly longer lengths of time (minutes to hours) and can only measure the size of DTNRs as a function of relative mobility to comigrating standards. Gel-based separation systems also suffer from a number of artifacts that reduce the accuracy of size measurements. These mobility artifacts are related to the specific sequences of DNA fragments and the persistence of secondary and tertiary structural elements even under highly denaturing conditions.

The inventors have performed significant work in developing time-of-flight mass spectrometry ("TOF-MS") as a means for separating and sizing DNA molecules, although other forms of mass spectrometry can be used and are within the scope of this invention. Balancing the throughput and high mass accuracy advantages of TOF-MS is the limited size range for which the accuracy and resolution necessary for characterizing DTNRs by mass spectrometry is available. Current state of the art for TOF-MS offers single nucleotide resolution up to ~100 nucleotides in size and four nucleotide resolution up to ~160 nucleotides in size. These numbers are expected to grow as new improvements are developed in the mass spectrometric field.

Existing gel-based protocols for the analysis of DTNRs do not work with TOF-MS because the allele PCR™ product size range, typically between 100 and 800 nucleotides, is outside the current resolution capabilities of TOF-MS. Application of DTNR analysis to TOF-MS requires the development of new primer sets that produce small PCR™ products 50 to 160 nucleotides in length, preferably 50 to 100 nucleotides in length. Amplified DNA may also be used to generate single stranded DNA products that are in the preferred size range for TOF-MS analysis by extending a primer in the presence of a chain termination reagent. A typical class of chain termination reagent commonly used by those of skill in the art is the dideoxynucleotide triphosphates. Again, application of DTNR analysis to TOF-MS requires that the primer be extended to generate products of 50 to 160 nucleotides in size, and preferably 50 to 100 nucleotides in length.

Gel-based systems are capable of multiplexing the analysis of 2 or more DTNR loci using two approaches. The first approach is to size partition the different PCR™ product loci. Size partitioning involves designing the PCR™ primers used to amplify different loci so that the allele PCR™ product size range for each locus covers a different and separable part of the gel size spectrum. As an example, the PCR™ primers for Locus A might be designed so that the allele size range is from 250 to 300 nucleotides, while the primers for Locus B are designed to produce an allele size range from 340 to 410 nucleotides.

The second approach to multiplexing 2 or more DTNR loci on gel-based systems is the use of spectroscopic partitioning. Current state of the art for gel-based systems involves the use of fluorescent dyes as specific spectroscopic markers for different PCR™ amplified loci. Different chromophores that emit light at different color wavelengths provide the means for differential detection of two different PCR™ products even if they are exactly the same size, thus 2 or more loci can produce PCR™ products with allele size ranges that overlap. For example, Locus A with a green fluorescent tag produces an allele size range from 250 to 300 nucleotides, while Locus B with a red fluorescent tag produces an allele size range of 270 to 330 nucleotides. A scanning, laser-excited fluorescence detection device monitors the wavelength of emissions and assigns different PCR™ product sizes, and their corresponding allele values, to their specific loci based on their fluorescent color.

In contrast, mass spectrometry directly detects the molecule preventing the use of optical spectroscopic partitioning as a means for multiplexing. While it is possible to have a limited use of size partitioning with TOF-MS, the limited size range of high-resolution detection by TOF-MS makes it likely that only 2 different loci can be multiplexed and size partitioned. In many cases, it may not be possible to even multiplex 2 loci and maintain a partitioning of the 2 different allele size ranges. Therefore, new methods are needed in order to employ mass spectrometry for the analysis of multiplexed DTNRs.

### SUMMARY OF THE INVENTION

It is, therefore, a goal of the present invention to provide newly designed PCR™ primers which are closer to the repeat regions than have previously been employed providing for the efficient analysis by TOF-MS. Specifically, the invention provides oligonucleotide primers designed to characterize various DTNR markers useful for human identity testing. The primers are for use in PCR™ amplification schemes, however, one of skill in the art could, in light of the present disclosure, employ them to generate appropriate size nucleic acid products for TOF-MS analysis using other methods of extending one or more of the disclosed primers. Additionally, these primers and their extension products are suitable for detection by mass spectrometry. Thus, applications of this invention include forensic and paternity testing and genetic mapping studies.

An embodiment of the present invention encompasses an oligonucleotide primer for use in analyzing alleles of a DNA tandem nucleotide repeat at a DNA tandem nucleotide repeat locus by mass spectrometry, which includes a nucleotide sequence that contains a flanking region of the locus where the primer upon extension generates a product that is capable of being analyzed by mass spectrometry. Preferably, the oligonucleotide primer's 3' end will be complementary to a region flanking a DNA tandem repeat region immediately adjacent to the DNA tandem repeat region or may further extend up to one, two, three, four or five tandem repeats into the DNA tandem repeat region. Used in this context "immediately adjacent" or "immediately flanking" means one, two, three, or four nucleotides away from the DNA tandem repeat region of the DNA tandem repeat locus.

The oligonucleotide primers of this invention are designed to generate extension products amenable to mass spectral analysis and containing a DTNR sequence, or region of interest, for which one is interested in determining the mass. The "flanking" regions of a DTNR locus are the portions of DNA sequence on either side of the DTNR region of interest. For embodiments employing PCR™ primers and polymerases to amplify a DTNR sequence, the primers are sufficiently

complementary to a portion of one or more flanking regions of the DTNR locus to allow the primer to effectively anneal to the target nucleic acid and provide a site to extend a complement to the target nucleic acid via PCR™. For embodiments employing primer extension, a preferred method is to use a single primer that is sufficiently complementary to allow effective annealing to a portion of a target DTNR locus flanking region in conjunction with a chain termination reagent. The chain termination reagent allows the production of discreet limited size nucleic acid products for mass spectral analysis. Preferred chain termination reagent for use in the present invention are dideoxynucleotide triphosphates. Therefore, for the methods comprising any type of primer extension, it is preferred that at least one of the primers is sufficiently complementary to a portion of a flanking region that is preferably adjacent to or close to the DTNR region of interest, generally within about 40 nucleotides of the DNA tandem nucleotide repeat region. As used in this context, "about" means anywhere from + 1 to 40 nucleotides, and all the integers in between, for example, +1, +2, +3, +4, +5, +6, +7, +8, +9, +10, etc. nucleotides.

The primer extension products are preferably single-stranded and may be any size that can be adequately resolved by mass spectrometric analysis. Preferably, detected, the final product single-stranded target nucleic acids are less than about 160 or 150 bases in length. More preferably, the extended nucleic acid products are from about 10 to 100 or 120 bases in length. As used in this context, "about" means anywhere from ± 1 to 20 bases, and all the integers in between, for example, +1, +2, +3, +4, +5, +6, +7, +8, +9, +10, etc. bases.

As used herein "a" will be understood to mean one or more. Thus, "a DNA tandem repeat marker" may refer, for example, to one, two, three, four, five or more DNA tandem repeat markers.

The present invention is also directed to new oligonucleotide primers which have been designed to match a portion of the flanking regions for various DTNR loci. Specific embodiments of this invention include oligonucleotide primers designed to amplify the following DTNR loci: CSF1PO, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, DYS19, F13A1, FES/FPS, FGA, HPRTB,



TH01, TPOX, DYS388, DYS391, DYS392, DYS393, D2S1391, D18S535, D2S1338, D19S433, D6S477, D1S518, D14S306, D22S684, F13B, CD4, D12S391, D10S220 and D7S523. With the exception of D3S1358, sequences for the STR loci of this invention are accessible to the general public through GenBank using the accession numbers listed in Table 1. These oligonucleotide primers may preferably contain a cleavable site, such as a recognition site for Type II and III restriction endonucleases, an exonuclease blocking site, or a chemically cleavable site, for reducing the length of the amplified product and increasing the mass spectral resolution.

Examples of some oligonucleotide primers that may be employed for amplifying these loci are listed in SEQ ID NO:1 through SEQ ID NO:103. Preferred oligonucleotide primers that also contain a cleavable phosphorothioate linkage and biotin moiety for immobilization on an avidin, streptavidin solid support are sequences according to SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100 and SEQ ID NO:103. These newly designed primers generate nucleic acid extension products which are smaller than those used previously with electrophoresis separation methods. Additionally, these primers may be used in other methods of primer extension known to those of skill in the art.

It will be apparent to one skilled in the art that some variations of these primers will also serve effectively, for example, adding or deleting one or a few bases from the primer and/or shifting the position of the primer relative to the DTR sequence by one or a few bases. Thus, primers encompassed by the present invention include the primers specifically listed as well as modifications of these primers. Although these sequences are all biotinylated at the 5' end and contain a phosphorothioate linkage at a particular location, one of skill in the art would

recognize that similar primers having biotin moieties and the cleavable groups at other sites would also be encompassed by the present invention. Primers containing types of immobilization attachments sites other than biotin, for example, would also be encompassed. Typically, the placement of the cleavable group is not critical as long as it is close enough to the 3' end to cleave the cleave the nucleic acid extension product to a reduced-length amplified product that is amenable to mass spectral analysis. These primers in pairs may also be combined to generate overlapping PCR™ product sizes which are all distinguishable by mass. However, for embodiments multiplexing multiple DTNR loci with overlapping allelic mass ranges, strategic placement of the cleavable group may effect a separation or an interleaving of mass spectral peaks.

Another embodiment of this invention encompasses a kit for analyzing alleles of a DTNR locus in a target nucleic acid, having a first strand and a second complementary strand, by mass spectrometry which includes a first primer complementary to the flanking region of a DNA tandem nucleotide repeat region and a second primer complementary to the opposite flanking region of a DNA tandem nucleotide repeat region. Preferred kits of this invention are kits for analyzing the following DTNR loci: CSF1PO, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, DYS19, F13A1, FES/FPS, FGA, HPR1B, TH01, TPOX, DYS388, DYS391, DYS392, DYS393, D2S1391, D18S535, D2S1338, D19S433, D6S477, D1S518, D14S306, D22S684, F13B, CD4, D12S391, D10S220 and D7S523.

Another embodiment of this invention encompasses a kit for analyzing alleles of a multiple DTNR loci in a target nucleic acid by mass spectrometry, which includes a plurality of primers complementary to the flanking regions of DNA tandem nucleotide repeat regions. Preferred kits of this invention are kits for analyzing the following DTNR loci: CSF1PO, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, DYS19, F13A1, FES/FPS, FGA, HPR1B, TH01, TPOX, DYS388, DYS391, DYS392, DYS393, D2S1391, D18S535, D2S1338,

D19S433, D6S477, D1S518, D14S306, D22S684, F13B, CD4, D12S391, D10S220 and D7S523.

The primers employed with these kits may preferably have cleavable sites, such as a recognition site for a restriction endonuclease, an exonuclease blocking site, or a chemically cleavable site. Preferred chemically cleavable sites encompass modified bases, modified sugars (e.g., ribose), and chemically cleavable groups incorporated into the phosphate backbone, such as dialkoxysilane, 3'-(S)-phosphorothioate, 5'-(S)-phosphorothioate, 3'-(N)-phosphoroamidate, or 5'-(N)-phosphoroamidate linkages. Another preferred embodiment is a kit employing a first primer that is capable of attaching to a solid support.

For primer extension by PCR amplification, it is preferable to employ these primers in pairs. Preferred pairs of primers include the following: a sequence according to SEQ ID NO:1 and a sequence according to SEQ ID NO:2; a sequence according to SEQ ID NO:3 and a sequence according to SEQ ID NO:4; a sequence according to SEQ ID NO:5 and a sequence according to SEQ ID NO:6; a sequence according to SEQ ID NO:7 and a sequence according to SEQ ID NO:8; a sequence according to SEQ ID NO:9 and a sequence according to SEQ ID NO:10; a sequence according to SEQ ID NO:11 and a sequence according to SEQ ID NO:12; a sequence according to SEQ ID NO:13 and a sequence according to SEQ ID NO:14; a sequence according to SEQ ID NO:15 and a sequence according to SEQ ID NO:16; a sequence according to SEQ ID NO:17 and a sequence according to SEQ ID NO:18; a sequence according to SEQ ID NO:19 and a sequence according to SEQ ID NO:20; a sequence according to SEQ ID NO:21 and a sequence according to SEQ ID NO:22; a sequence according to SEQ ID NO:23 and a sequence according to SEQ ID NO:24; a sequence according to SEQ ID NO:25 and a sequence according to SEQ ID NO:26; a sequence according to SEQ ID NO:27 and a sequence according to SEQ ID NO:28; a sequence according to SEQ ID NO:29 and a sequence according to SEQ ID NO:30; a sequence according to SEQ ID NO:31 and a sequence according to SEQ ID NO:32; a sequence according to SEQ ID NO:49 and a sequence according to SEQ ID NO:83; a sequence according to SEQ ID NO:52 and a sequence according to SEQ ID NO:84; a sequence

according to SEQ ID NO:54 and a sequence according to SEQ ID NO:85; a sequence according to SEQ ID NO:56 and a sequence according to SEQ ID NO:86; a sequence according to SEQ ID NO:58 and a sequence according to SEQ ID NO:87; a sequence according to SEQ ID NO:59 and a sequence according to SEQ ID NO:88; a sequence according to SEQ ID NO:62 and a sequence according to SEQ ID NO:89; a sequence according to SEQ ID NO:63 and a sequence according to SEQ ID NO:90; a sequence according to SEQ ID NO:66 and a sequence according to SEQ ID NO:91; a sequence according to SEQ ID NO:67 and a sequence according to SEQ ID NO:92; a sequence according to SEQ ID NO:70 and a sequence according to SEQ ID NO:93; a sequence according to SEQ ID NO:72 and a sequence according to SEQ ID NO:94; a sequence according to SEQ ID NO:74 and a sequence according to SEQ ID NO:95; a sequence according to SEQ ID NO:76 and a sequence according to SEQ ID NO:96; a sequence according to SEQ ID NO:78 and a sequence according to SEQ ID NO:97; a sequence according to SEQ ID NO:80 and a sequence according to SEQ ID NO:98; a sequence according to SEQ ID NO:66 and a sequence according to SEQ ID NO:99; a sequence according to SEQ ID NO:33 and a sequence according to SEQ ID NO:100; and a sequence according to SEQ ID NO:101 and a sequence according to SEQ ID NO:103.

In one embodiment, at least one of the primers used to prepare the nucleic acid extension product contains a surface binding moiety, such as a biotin moiety, at the 5'-end and a cleavable moiety, such as a phosphorothioate linkage (see FIGS. 7A and 7B), near the 3'-end for a capture and release assay, such as one using streptavidin-coated magnetic beads for binding biotinylated primers, described in PCT Patent Application No. WO 96/37630, and incorporated herein by reference. These linkages are often referred as thiophosphate linkages as well. Incorporation of a method for obtaining single-stranded PCR™ products, such as is possible with the primer modifications described above, is preferred. Removal of one of the two strands halves the number of DNA oligomers that will be visualized by TOF-MS and improves the likelihood of resolving all PCR™ product strands.

Another embodiment of this invention encompasses a method for analyzing DNA tandem nucleotide repeat alleles at a DNA tandem nucleotide repeat locus in a

target nucleic acid by mass spectrometry which includes the steps of a) obtaining a target nucleic acid containing a DNA tandem nucleotide repeat region; b) extending the target nucleic acid using one or more primers to obtain a limited size range of nucleic acid extension products, wherein the primers are complementary to a sequence  
5 flanking the DNA tandem nucleotide repeat of said locus; and c) determining the mass of the nucleic acid extension products by mass spectrometry, where the target nucleic acid is normally double-stranded (i.e. it has a first strand and a second complementary strand). Nucleic acid extension products may be generated in this method by any means known to those of skill in the art, and particularly either by amplification, such  
10 as PCR amplification, or by primer extension in conjunction with a chain termination reagent. Preferred primers may immediately flank the DNA tandem repeat locus, or may further extend up to one, two, three, four or five tandem repeats into the DNA tandem repeat region. Used in this context "immediately adjacent" or "immediately flanking" means one, two, three, or four nucleotides away from the DNA tandem  
15 repeat region of the DNA tandem repeat locus. Preferred primers may contain a cleavable site, such as a recognition site for a restriction endonuclease, an exonuclease blocking site, or a chemically cleavable site, and be capable of attaching to a solid support.

These primers may be capable of directly or indirectly attaching to a solid  
20 support via covalent or noncovalent binding. The primers may contain an immobilization attachment site (IAS) for attachment to a solid support. This site is usually upstream of the chemically cleavable site. A suitable immobilization attachment site is any site capable of being attached to a group on a solid support. These sites may be a substituent on a base or sugar of the primer. An IAS may be, for  
25 example, an antigen, biotin, or digoxigenin. This attachment allows for isolation of only one strand of an amplified product. Such isolation of either single-stranded or double-stranded amplified target nucleic acids generally occurs prior to the application of the nucleic acids to the matrix solution, resulting in well-defined mass spectral peaks and enhanced mass accuracy. The matrix solution can be any of the  
30 known matrix solutions used for mass spectrometric analysis, including 3-

hydroxypicolinic acid ("3-HPA"), nicotinic acid, picolinic acid, 2,5-dihydroxybenzoic acid, and nitrophenol.

For example, in one embodiment, a strand of a target nucleic acid extension product may be bound or attached to a solid support to permit rigorous washing and concomitant removal of salt adducts, unwanted oligonucleotides and enzymes. Either  
5 a double-stranded or a single-stranded nucleic acid extension product may be isolated for mass spectrometric analysis. The single-stranded target nucleic acid extension product analyzed by MS may be either the strand bound or not bound to the solid support.

10 When an unbound strand is used for MS analysis, it is typically purified by first washing the bound strand and its attached complement under conditions not sufficiently rigorous to disrupt the strand's attachment to its bound complement. After unwanted biomolecules and salts are removed, the complement may then be released under more rigorous conditions. In contrast, when the bound strand is to be  
15 analyzed, it is typically washed under more vigorous conditions such that the interactions between the bound strand, if present, and its unbound complement is disrupted. This allows the unbound strand to be washed away with the other salts and unwanted biomolecules. Cleavable linkers or cleavable primers may be used to release the bound strand from the solid support prior to MS analysis.

20 Preferred primers for practicing this method include primers designed to amplify DTNR loci selected from the group consisting of CSF1PO, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, DYS19, F13A1, FES/FPS, FGA, IIPRTB, TH01, TPOX, DYS388, DYS391, DYS392, DYS393, D2S1391, D18S535, D2S1338, D19S433, D6S477, D1S518, D14S306, D22S684,  
25 F13B, CD4, D12S391, D10S220 and D7S523. Preferred pairs of primers designed to amplify these loci include: a sequence according to SEQ ID NO:1 and a sequence according to SEQ ID NO:2; a sequence according to SEQ ID NO:3 and a sequence according to SEQ ID NO:4; a sequence according to SEQ ID NO:5 and a sequence according to SEQ ID NO:6; a sequence according to SEQ ID NO:7 and a sequence  
30 according to SEQ ID NO:8; a sequence according to SEQ ID NO:9 and a sequence

according to SEQ ID NO:10; a sequence according to SEQ ID NO:11 and a sequence  
according to SEQ ID NO:12; a sequence according to SEQ ID NO:13 and a sequence  
according to SEQ ID NO:14; a sequence according to SEQ ID NO:15 and a sequence  
according to SEQ ID NO:16; a sequence according to SEQ ID NO:17 and a sequence  
5 according to SEQ ID NO:18; a sequence according to SEQ ID NO:19 and a sequence  
according to SEQ ID NO:20; a sequence according to SEQ ID NO:21 and a sequence  
according to SEQ ID NO:22; a sequence according to SEQ ID NO:23 and a sequence  
according to SEQ ID NO:24; a sequence according to SEQ ID NO:25 and a sequence  
according to SEQ ID NO:26; a sequence according to SEQ ID NO:27 and a sequence  
10 according to SEQ ID NO:28; a sequence according to SEQ ID NO:29 and a sequence  
according to SEQ ID NO:30; a sequence according to SEQ ID NO:31 and a sequence  
according to SEQ ID NO:32; a sequence according to SEQ ID NO:49 and a sequence  
according to SEQ ID NO:83; a sequence according to SEQ ID NO:52 and a sequence  
according to SEQ ID NO:84; a sequence according to SEQ ID NO:54 and a sequence  
15 according to SEQ ID NO:85; a sequence according to SEQ ID NO:56 and a sequence  
according to SEQ ID NO:86; a sequence according to SEQ ID NO:58 and a sequence  
according to SEQ ID NO:87; a sequence according to SEQ ID NO:59 and a sequence  
according to SEQ ID NO:88; a sequence according to SEQ ID NO:62 and a sequence  
according to SEQ ID NO:89; a sequence according to SEQ ID NO:63 and a sequence  
20 according to SEQ ID NO:90; a sequence according to SEQ ID NO:66 and a sequence  
according to SEQ ID NO:91; a sequence according to SEQ ID NO:67 and a sequence  
according to SEQ ID NO:92; a sequence according to SEQ ID NO:70 and a sequence  
according to SEQ ID NO:93; a sequence according to SEQ ID NO:72 and a sequence  
according to SEQ ID NO:94; a sequence according to SEQ ID NO:74 and a sequence  
25 according to SEQ ID NO:95; a sequence according to SEQ ID NO:76 and a sequence  
according to SEQ ID NO:96; a sequence according to SEQ ID NO:78 and a sequence  
according to SEQ ID NO:97; a sequence according to SEQ ID NO:80 and a sequence  
according to SEQ ID NO:98; a sequence according to SEQ ID NO:66 and a sequence  
according to SEQ ID NO:99; a sequence according to SEQ ID NO:33 and a sequence

according to SEQ ID NO:100; and a sequence according to SEQ ID NO:101 and a sequence according to SEQ ID NO:103.

The present invention also focuses on an improved method of multiplexing the analysis of nucleic acid extension products derived from DNA nucleotide repeat loci.

5 This method differs from known methods of multiplexing DTNR analysis in that mass spectrometry is employed and the range of possible nucleic acid extension products for the multiplexed loci, the allele nucleic acid extension product size ranges, may be specifically chosen to overlap in the mass scale yet be uniquely resolved and detected.

Thus, this invention encompasses methods for analyzing more than one target  
10 nucleic acid in which the target nucleic acids are used to produce more than one nucleic acid product extension product and where each nucleic acid extension product may comprise a different DTNR sequence. A preferred embodiment encompasses simultaneously determining the mass of more than one DNA tandem nucleotide repeat allele at more than one DNA tandem nucleotide repeat loci. According to this  
15 embodiment several amplification products containing various DTNR sequences from different DTNR loci may be analyzed in the same solution and spectrum.

Additionally, the DNA tandem nucleotide repeat loci may have overlapping allelic mass ranges (see FIGS. 4 and 5). The term "overlapping allelic mass ranges" is defined to mean that the alleles that may be present for a particular DTNR locus have  
20 masses that overlap, or coincide, as observed by mass spectrometry with the masses for alleles from another DTNR locus. The methods of the present invention allow one to resolve these alleles by mass spectrometry either by increasing the mass separation of these peaks or by modifying the mass of the amplified products containing the various DTNR sequences such that the amplification products have interleaving mass  
25 spectral peaks (see FIG. 6).

This novel interleaved multiplexing approach overcomes the TOF-MS limitations for size partitioning and takes advantage of the high mass accuracy of the method within the high resolution mass range below about 160 nucleotides in size. One specific embodiment encompasses a method that involves the design of specific  
30 primer or primers that produce nucleic acid extension products for a first locus with



defined allele mass values. The primer or primers for second locus are then selected so that while the mass range for the predicted nucleic acid extension products of the primers overlap with the mass range for the products of the first locus, the specific predicted nucleic acid extension product mass values differ from those of the first locus and therefore can be uniquely resolved by TOF-MS. Further loci may be added to the multiplex using the same method such that three, four, five, six, seven, eight, nine, ten or more loci may be analyzed simultaneously.

The basic limits for this multiplexing are defined by the ability to resolve all possible nucleic acid extension products within a mixture. It is not inconceivable that as many as 10 different loci might be interleaved and uniquely resolved. In addition to multiplexing two or more DTNRs it is also possible to use this invention to interleave mixtures of DTNRs with specific nucleic acid extension products arising from nonrepeat loci, e.g., a DTNR locus with allelic nucleic acid extension products 72, 76, 80, 84 and 88 nucleotides in size could be simultaneously analyzed with a nucleic acid extension product 82 nucleotides in size.

The ability to interleave loci requires that the nucleic acid extension product mass values for all possible allele values should preferably be known. These allele mass values may be determined empirically or more likely by calculation using the known loci sequences. In many cases it may be necessary to "fine tune" the allele mass values for one or more loci in a multiplexed mixture in order to prevent unresolvable overlap between two Nucleic acid extension products. For example, allele 5 for Locus A may be only 5 Da different in mass than allele 9 for Locus B preventing resolution of those two Nucleic acid extension products by mass spectrometry. Mass modifications to one or both loci may be used to increase this mass difference to 100 Da.

Adjusting the allele mass values for any given locus may be done by any number of methods including: increasing or decreasing the size of the nucleic acid extension products via altered sequences and placement of the primers; addition of nonhybridizing nucleotides to the 5' ends of one or more primers; addition of nonnucleotide chemical modifications internally or to the ends of one or both primers;

alterations in base composition within one or both primers, including the use of nonstandard nucleotides, that may or may not result in mismatches within the primers; incorporation of and specific placement of a chemically cleavable moiety within the primer backbone to reduce the length of the nucleic acid extension product by a selected amount; enzymatic cleavage of the nucleic acid extension products using a restriction endonuclease that recognizes a restriction site within one or both primers or within the nucleic acid extension product itself; use of a 5' to 3' exonuclease in concert with exonuclease blocking modified nucleotides contained within one or more primers; incorporation of nonstandard deoxyribonucleotides or chemically or isotopically modified nucleotides during polymerization; any number of methods of mass modifying by addition of chemical moieties post amplification; by using different chain termination reagents in conjunction with primer extension; or any number of other means that anyone skilled in the art would be able to identify.

Another embodiment encompasses a method of multiplexing amplification products containing DTNRs having overlapping allelic ranges where at least one amplification product contains a mass modified nucleotide. Mass modified nucleotides include nucleotides to which nonnucleotide moieties have been chemically attached; bases having altered compositions; nonstandard nucleotides, that may or may not result in mismatches within the primers; and any bases whose masses have been modified through the addition of chemical moieties after the amplification step.

Alternatively, the length of at least one extension product may be reduced by cleaving the extension product at a cleavable site such as a restriction endonuclease cleavage site, an exonuclease blocking site, or a chemically cleavable site. Preferred chemically cleavable sites for multiplexing include modified bases, modified sugars (e.g., ribose), or a chemically cleavable group incorporated into the phosphate backbone, such as a dialkoxysilane, 3'-(S)-phosphorothioate, 5'-(S)-phosphorothioate, 3'-(N)-phosphoramidate, or 5'-(N)-phosphoramidate. Preferred primers may also be capable of attaching to a solid support.

Another embodiment of this invention encompasses a method for multiplexing the detection of more than one amplified DNA tandem nucleotide repeat marker from more than one DNA tandem nucleotide repeat loci including: determining the mass of more than one nucleic acid extension product by mass spectrometry, where the DNA  
5 tandem nucleotide repeat loci each comprise a DNA tandem repeat sequence and a flanking sequence and have overlapping allelic mass ranges. Typically, at least one of the target nucleic acid extension products may contain a mass modifying group.

"Mass modifying groups" may comprise any group that alters the mass of the amplified products to produce interleaving or otherwise resolvable mass spectral  
10 peaks. These groups, which may be incorporated during or after primer extension, may be mass modified nucleotides, nonstandard deoxyribonucleotides, or even cleavable sites as cleaving such a site modifies the mass by reducing the length of the extension product. As used in this context, modified or nonstandard bases are generally understood to include bases not found in DTNR locus flanking the DTNR  
15 sequence of the sample or target nucleic acid.

#### **BRIEF DESCRIPTION OF THE DRAWINGS**

FIG. 1 is a mass spectrum of an allelic ladder from the tyrosine hydroxylase gene ("TH01"). Most of the common alleles for this STR marker (alleles 5, 6, 7, 8, 9,  
20 9.3, and 10) can be seen. Alleles 9.3 and 10 differ by a single nucleotide while the other alleles are separated by four bases.

FIG. 2 displays mass spectra for several samples from the TPOX locus. The top spectrum is an allelic ladder containing alleles ranging from 6 to 13 repeats while  
25 the other spectra show the isolation of various alleles for this locus.

FIG. 3A displays the mass spectrum for the CSF1PO locus.

FIG. 3B displays the mass spectrum for the D3S1358 locus.

FIG. 3C displays the mass spectrum for the D5S818 locus.

FIG. 3D displays the mass spectrum for the D7S820 locus.

30 FIG. 3E displays the mass spectrum for the D8S1179 locus.

**FIG. 3F** displays the mass spectrum for the D13S317 locus.

**FIG. 3G** displays the mass spectrum for the D16S539 locus.

**FIG. 3H** displays the mass spectrum for the D18S51 locus.

**FIG. 3I** displays the mass spectrum for the D21S11 locus.

5 **FIG. 3J** displays the mass spectrum for the DYS19 locus.

**FIG. 3K** displays the mass spectrum for the F13A1 locus.

**FIG. 3L** displays the mass spectrum for the FES/FPS locus.

**FIG. 3M** displays the mass spectrum for the FGA locus.

**FIG. 3N** displays the mass spectrum for the HPR1B locus.

10 **FIG. 3O** displays the mass spectrum for the TH01 locus.

**FIG. 3P** displays the mass spectrum for the TPOX locus.

**FIG. 4** is a simulated multiplex STR analysis of alleles with overlapping size ranges. This diagram depicts the expected masses for known alleles of TPOX and TH01.

15 **FIG. 5** are mass spectra of mixtures of TH01 and TPOX allelic ladders. Using the primer sequences for TH01 (SEQ ID NO.:29 and SEQ ID NO.:30) and TPOX (SEQ ID NO.:31 and SEQ ID NO.:32), alleles between the different STR systems differ by only 120 Daltons (top spectrum). By adding two nucleotides to the 5'-end of the reverse primer for TPOX (SEQ ID NO.:32), the TPOX allele masses are increased  
20 by ~600 Daltons, making them easier to resolve.

**FIG. 6** is a simulated multiplex STR analysis depicting the expected masses for D16S539 and D7S820 known alleles. Even with different repeat sequences, all known alleles may be resolved by mass spectroscopy.

25 **FIG. 7A** shows the chemical formula for 2'-deoxythymidine-3'-(S)-phosphorothioate.

**FIG. 7B** shows the chemical formula for 2'-deoxythymidine-5'-(S)-phosphorothioate.

**FIG. 8A** shows the expected allele sizes for CTT multiplex analyses. The CTT multiplex is directed to the three STR loci CSF1PO, TPOX, and TH01.

FIG. 8B illustrates the results of the analysis of a sample using the CTT multiplex. The sample is shown to contain a homozygous TPOX allele 8, heterozygous TH01 alleles 6 and 9.3, and a homozygous CSF1PO allele 12.

## 5                    DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

The present invention focuses on a mass spectrometric method of multiplexing the analysis of Nucleic acid extension products which overlap in mass derived from DNA nucleotide repeat loci. For example, to resolve all possible alleles of the  
10    DTNRs being analyzed the masses of the Nucleic acid extension products from two or more DTNR markers may be offset from one another so that any two possible alleles (or any two possible common alleles) do not overlap in mass within the mass resolution of the mass spectrometer, yet the ranges of the possible alleles do overlap. Within the overlapping mass range, defined as the mass range held in common by two  
15    loci with defined allele size ranges, the DTNR marker may be offset from one another by some fraction of the mass of the sequence repeat unit, e.g. for tetranucleotide DTNR markers mass offsets less than four nucleotides, for dinucleotide DTNRs mass offsets less than 2 nucleotides. Other types of offset, such as may be found when multiplexing dinucleotide repeat loci with tetranucleotide or complex nucleotide  
20    repeat loci, will be apparent to one skilled in the art.

This approach overcomes the TOF-MS limitations for size partitioning, where the PCR™ product for the allele range of two or more sets of possible loci do not overlap, by taking advantage of the high mass accuracy associated with mass spectroscopy within the high resolution mass range (below ~160 nucleotides in size).  
25    Although this method is currently most useful for oligonucleotides below ~160 nucleotides, this size is a function of the number of nucleotides in the repeat as well as the resolution of the mass spectroscopic method. Therefore, larger oligonucleotides are also useful with the present invention, particularly where larger repeat sequences (tetra- vs. dinucleotides) or as advances in mass spectroscopy allow for greater mass  
30    resolution in higher mass ranges.

This multiplexing method involves the design of specific primers that produce Nucleic acid extension products for a first locus with defined allelic mass values. The primers for the second locus are then chosen so that while the mass range for the different alleles overlaps with the mass range for the first locus, the specific allele mass values differ from those of the first locus and therefore can be uniquely resolved by TOF-MS. The identity of each allele, defined by the specific Nucleic acid extension products being characterized, is uniquely determined using the high accuracy molecular mass values provided by TOF-MS. In contrast, gel-based methods are not capable of providing accurate mass values for uniquely identifying each product within a multiplexed, allelically interleaved mixture of Nucleic acid extension products. The basic limits for this multiplexing method are defined by the ability to resolve all possible, or all common, Nucleic acid extension products within a mixture. Potentially as many as 10 different loci might be interleaved and fully resolved.

The invention further relates to primers designed to characterize 33 DNA repeat markers useful for human identity testing. Applications include forensic and paternity testing as well as genetic mapping studies. These D1NR markers are useful in PCR<sup>™</sup> amplification, preferably as pairs of oligonucleotide primers, and in other methods of primer extension may be used as single primers, the extension products of which may be accurately detected by mass spectrometry as they are smaller than those used previously with electrophoresis separation methods.

These new oligonucleotide primers are designed to match a portion of the flanking regions for D1NR loci consisting of: CSF1PO, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, DYS19, F13A1, FES/FPS, FGA, HPR1B, TH01, TPOX, DYS388, DYS391, DYS392, DYS393, D2S1391, D18S535, D2S1338, D19S433, D6S477, D1S518, D14S306, D22S684, F13B, CD4, D12S391, D10S220 and D7S523. With the exception of D3S1358, sequences for the STR loci of this invention are accessible to the general public through GenBank using the accession numbers listed in Table 1. The sequence ID Numbers given in Table 1 correspond to the DNA sequence of the DNA tandem repeat regions of each locus and

its flanking regions. Flanking sequences further from the DTNR region could easily be obtained by one of skill in the art by accessing the GenBank listings. FIGS. 3A-3P display mass spectra for each of the STR loci listed in TABLE 1. It will be apparent to one skilled in the art that small variation of these primers will also serve effectively, for example, adding or deleting one or a few bases from the primer and/or shifting the position relative to the template sequence by one or a few bases.

The use of a hybridization probe of about 14-25 nucleotides in length allows the formation of a duplex molecule that is both stable and selective. Molecules having contiguous complementary sequences over stretches greater than 14 bases in length are generally preferred, though, in order to increase stability and selectivity of the hybrid, and thereby improve the quality and degree of specific hybrid molecules obtained. One will generally prefer to design nucleic acid molecules having gene-complementary stretches of 15 to 25 contiguous nucleotides, or even longer where desired.

Hybridization probes may be selected from any portion of any of the sequences disclosed herein. All that is required is to review the primer sequences set forth in Table I or to any continuous portion of the sequence as in the DTNR loci, whose locus sequence ID numbers are listed in Table 1 or any other DTNR locus, from about 14-25 nucleotides in length up to and including the full length sequence, that one wishes to utilize as a probe or primer. The choice of probe and primer sequences may be governed by various factors known to those of skill in the art.

The process of selecting and preparing a nucleic acid segment that includes a contiguous sequence from within the DTNR loci, whose locus sequence ID numbers are listed in Table 1 or any other DTNR locus, may alternatively be described as preparing a nucleic acid fragment. Of course, fragments may also be obtained by other techniques such as, e.g., by mechanical shearing or by restriction enzyme digestion. Small nucleic acid segments or fragments may be readily prepared by, for example, directly synthesizing the fragment by chemical means, as is commonly practiced using an automated oligonucleotide synthesizer. Also, fragments may be obtained by application of nucleic acid reproduction technology, such as the PCR™

technology of U. S. Patent 4,683,202 (incorporated herein by reference), by introducing selected sequences into recombinant vectors for recombinant production, and by other recombinant DNA techniques generally known to those of skill in the art of molecular biology.

- 5           Accordingly, the nucleotide sequences of the invention may be chosen for their ability to selectively form duplex molecules with complementary stretches of the flanking regions of DNA nucleotide repeat regions. Depending on the application envisioned, one will desire to employ varying conditions of hybridization to achieve  
10           varying degrees of selectivity of probe towards target sequence. For applications requiring high selectivity, one will typically desire to employ relatively stringent conditions to form the hybrids, e.g., one will select relatively low salt and/or high temperature conditions, such as provided by a salt concentration of from about 0.02 M to about 0.15 M salt at temperatures of from about 50°C to about 70°C. Such selective conditions tolerate little, if any, mismatch between the probe and the  
15           template or target strand.

- Of course, for some applications, less stringent (reduced stringency) hybridization conditions will be tolerated by the primer extension system in order to allow sufficiently specific formation of the heteroduplex of primer and target DNA. In these circumstances, one may desire to employ salt conditions such as those of  
20           from about 0.15 M to about 0.9 M salt, at temperatures ranging from about 20°C to about 55°C. Cross-hybridizing species can thereby be readily identified as positively hybridizing signals with respect to control hybridizations. In any case, it is generally appreciated that conditions can be rendered more stringent by the addition of increasing amounts of formamide, which serves to destabilize the hybrid duplex in the  
25           same manner as increased temperature. Thus, hybridization conditions can be readily manipulated to ensure that a primer sequence will yield extension product mainly from the desired target D1NR locus.



TABLE 1

PrimerSEQ ID NO. <sup>1</sup>	Primer Sequence (5'—3')	STR Locus <sup>2</sup>	Locus SEQ ID No. <sup>3</sup> (GenBank Accession)	PCR <sup>TM</sup> Size <sup>4</sup>	Repeat <sup>5</sup>
1, 100	ACAGTAACIGCCCTTCATAGATAG	CSF1PO-F	104	12=113bp	AGAT
2, 33	GTGTCAGACCCCTGTTCTAAGTA	CSF1PO-R	(X14720)		
3	ACTGCAGTCCCAATCTGGGT	D3S1358-F	---	16-109bp	GAYA
4, 34	ATGAAATCAACAGAGGCTTG	D3S1358-R	---		
5, 35	CTCTTTGGTATCCTTATGTAATATT	D5S818-F	105	11=105bp	AGAT
6	ATCTGTATCCTTATTTATACCTCTATCTA	D5S818-R	(G08446)		
7, 36	TGTCATAGTTTAGAACGAACTAAC	D7S820-F	106	12=90bp	GATA
8	GAAAAACTATCAATCTGTCTATCTAT	D7S820-R	(G08616)		
9, 37	TTTGTAATTCATGTCTACATTCTGATC	D8S1179-F	107	12=106bp	TATC
10	ACCTATCCTGTGAGATTAATTTTACATGTG	D8S1179-R	(G08710)		
11, 38	CCCATCTAAGCCCTATCTGTATTT	D13S317-F	108	13=122bp	TATC
12	GCCCCAAAAGACAGACAGACAGAAAG	D13S317-R	(G09017)		
13	AGACAGACAGACAGGTGGATAGA	D16S539-F	109	11=83bp	GATA
14, 39	TCTCTGTTTGTCTTTTCAATGATA	D16S539-R	(G07925)		

TABLE 1 (continued)

PrimerSEQ	Primer Sequence (5'—3')	STR	Locus <sup>3</sup>	Locus SEQ ID No. <sup>3</sup>	PCR <sup>2a</sup> Size <sup>4</sup>	Refseq <sup>5</sup>
ID NO. <sup>1</sup>				(GenBank Accession)		
15	TGAGTCACAAATTGAGACCTT	D18S11-F		110	13-144bp	AGAA
16, 40	GTCCTACAAATAACAGTTGCTACTATT	D18S11-R		(L18333)		
17, 41	CCCAAGTGAATTGOCITCTA	D21S11-F		111	26-150bp	TCTR
18	GTAGATAGACTGGATAGATAGACGATAG	D21S11-R		(M84567)		
19, 42	GTGTTTTAGATAGATAGATAGGTA	DYS19-F		112	10-84bp	TAGA
20	GGTTAAGGAGAGTGTCACCTA	DYS19-R		(X77751)		
21, 43	CAGAGCAAGACTTCACTCIG	F13A1-F		113	7-128bp	AAAG
22	TCAATTTTAGTGCATGTTTC	F13A1-R		(M21986)		
23, 44	TTAGGAGACAAGGATAGCAGTTC	FES/TPS-F		114	11-91bp	ATTI
24	CCGAAAGAAATGAGACTACATCT	FES/TPS-R		(X06292)		
25, 45	AAAATTAGGGCATATTACAAAGCTAGTT	FGA-F		115	21-142bp	CTTT
26	TCTGTAAATTGCCAGCAAAAAGAAA	FGA-R		(M64982)		
27, 46	GTCTCCATCTTTGTCTCTATCTCTATCTG	HPR1B-F		116	13-108bp	TCTA
28	GAGAAAGGCAATGAATTTCCTT	HPR1B-R		(M26434)		

TABLE 1 (continued)

PrimerSEQ	Primer Sequence (5' → 3')	STR	Locus SEQ ID No. <sup>3</sup>	PCR <sup>1M</sup>	Repeat <sup>5</sup>
ID NO. <sup>1</sup>		Locus <sup>2</sup>	(GenBank Accession)	Size <sup>4</sup>	
29	CCTGTTCTCTCCCTTATTCCC	TI101-F	117	9=79bp	TCAAT
30, 47	GGGAACACAGACTCCATGGT	TI101-R	(D00269)		
31, 48	CTTAGGGAAACCCTCACAGATG	TPOX-F	118	11=89bp	AATG
32	GTCCCTTGTCAGCGTTTATTTC	TPOX-R	(M68651)		
49	GTGAGTTAGCCGTTTAGCGAT	DYS38B-F	119	17=108bp	ATT
50, 83	GAGCGAGAGTCCGTCTCA	DYS38B-R	(G09695)		
51, 84	TTCAAATCATACACCCATATCTGTC	DYS391-F	120	9=99bp	TCTR
52	ATAGAGGGATAGGTAGGCAGGC	DYS391-R	G09613		
53, 85	TTTTCTTGATCACCATT	DYS392-F	121	16=98bp	TAT
54	AAACCTACCAATCCCAATCCCT	DYS392-R	G09867		
55, 86	TGGCTCTCTACCTGTGTCAATAC	DYS393-F	122	15=106bp	AGAT
56	TGTCTCATAGAAAAGACATACAT	DYS393-R	G09601		
57, 87	CTGGATTCTTGGTTATAGTAAA	D2S1391-F	123	12=100bp	TCTA
58	AAGCTGTTAGAGAGATACACAGA	D2S1391-R	G08168		

TABLE 1 (continued)

Primer SEQ ID NO. <sup>1</sup>	Primer Sequence (5'—3')	STR Locus <sup>2</sup>	Locus SEQ ID No. <sup>3</sup>	PCR Size <sup>4</sup>	Repeat
59	AGCCACACCCATAAAGTTT	D18S33-F	124	14=120bp	GATA
60, 88	GAATGCAGAGAAAGAGAAATCTA	D18S33-R	G07985		
61, 89	AGAAAAGGCTTGGCCCTTG	D2S1338-F	125	11=100bp	CCTT
62	TAAAGGATTCACAGGAGGG	D2S1338-R	G08202		
63	GAATAAGATTCTGTGGAAGGAAA	D19S433-F	126	11=100bp	AACG
64, 90	AATCTTCTCTCTTTTCTACCTCTCT	D19S433-R	G08036		
65, 91	AGGCTGATGAGGTGAAATA	D6S477-F	127	16=120bp	ATCT
66	ICAAACAACAACACATAAAGATGA	D6S477-R	G08543		
67	CATATATTTGTAGATGGATAGAAGA	D1S518-F	128	14=105bp	GATA
68, 92	GAGTTCTCCAGAGAAACAGAAATC	D1S518-R	G07854		
69, 93	CAGACTAGATAGATAGATACGTACATACA	D14S306-F	129	14=139bp	AGAT
70	TCAAAGAGTGACAAAGAACTAAA	D14S306-R	G09055		
71, 94	CCATCCATCTATCATCTATTATT	D22S684-F	130	11=100bp	TATC
72	ACCTACATTAGTCTGTGTTCTCT	D22S684-R	G08089		

TABLE 1 (continued)

PrimerSEQ ID NO. <sup>1</sup>	Primer Sequence (5'—3')	STR Locus <sup>2</sup>	Locus SEQ ID No. <sup>3</sup> (GenBank Accession)	PCR <sup>4</sup> Size <sup>1</sup>	Repeat <sup>5</sup>
73, 95	AAGAAAGAAATGACCCCTTGGAAATTT	F13B-F	131	10=97bp	TTTA
74	GGGCGACAGAGCAAGACTC	F13B-R	M64554		
75, 96	TGGAGTCGCAAGCTGAACTA	CD4-F	132	9-108bp	TTTTC
76	CTGAGTGACACAGGTGAGAACCTG	CD4-R	M86525		
77, 97	ATCAATGGATGCATAGGTA	D12S391-F	133	20=142bp	YAGA
78	GCCTCCATATCACTTGAGCTAAT	D12S391-R	G08921		
79, 98	GCCTTACTGACTTACTACATAACGA	D10S220-F	134	23=100bp	CA
80	GAGCAAGACTGCCATCTCAAAA	D10S220-R	Z17087		
81, 99	TGGAAAAATAATCTGGGAAGATA	D7S523-F	135	17=100bp	CA
66	CCTGTTGACATTTTAAACCA	D7S523-R	Z17102		
101	GCCGTGTCCTCCCTTATTTCCT	TH01-F	117	9-88bp	TCAAT
102, 103	AGGTCACAGGGAACACAGACTCC	TH01-R	D00269		

<sup>1</sup>Bold sequence numbers correspond to primer sequences containing sequence modifications including biotinylation and the presence of a cleavable phosphorothioate linkage.

<sup>2</sup>F and R indicate forward and reverse primers for each locus.

<sup>3</sup>The sequence listings contain the Genbank sequence for each of the tandem repeat loci including the DNA tandem repeat region and flanking regions for each locus. The sequence listings correspond to only a portion of the full Genbank sequence listing.

<sup>4</sup>The first number in the PCR product size is the number of repeats found in the Genbank sequence listing for each locus and the second is the predicted size of PCR product from the Genbank sequence when using the listed primers to amplify the tandem repeat locus. Of course, the number of tandem repeats within a population of individuals will vary and therefore so will the PCR product size when individual samples are analyzed.

<sup>5</sup>Repeats sequence nomenclature used here is according to the latest recommendations of the DNA Commission of the International Society for Forensic Haemogenetics, as described in *Int. J. Legal Med.* 110:175-176 (1997).

At least one of the primers in each locus-specific pair contains a biotin moiety at the 5'-end and a phosphorothioate linkage attached to a T near the 3'-end for a capture and release assay using streptavidin-coated magnetic beads (PCT Patent Application No. WO 96/37630). Although many of the specific primers of the present invention are designed for use in such a capture and release assay, such primers need not contain either solid-binding or cleavable sites, or may contain any combination of them.

The purpose of such an assay is to increase mass resolution by (1) allowing for higher purities of the nucleic acid extension product and (2) decreasing the size of the nucleic acid extension product. Binding to a solid support fulfills the first goal by allowing for stringent washes and removing the complementary strand of the DNA which provides cumulative information and complicates the mass spectra decreasing the resolution.

This assay may be used to isolate single-stranded or double-stranded amplified target nucleic acids. Typically, at least one strand of an amplified target nucleic acid is bound to a solid support to permit rigorous washing and concomitant removal of salt adducts, unwanted oligonucleotides and enzymes. Either a double-stranded amplified target nucleic acid or a single-stranded amplified target nucleic acid may be isolated for mass spectrometric analysis depending upon the stringency of the wash. The single-stranded amplified target nucleic acid analyzed may be either the strand bound or not bound to the solid support. If the unbound strand is used for MS analysis, it is purified by first washing the bound strand and its attached complement under conditions not sufficiently rigorous to disrupt the strand. After unwanted biomolecules and salts are removed, the complement can then be released under more rigorous conditions. Cleavable linkers or cleavable primers may then be used to release the bound strands from the solid support prior to MS analysis.

The second goal is met by having cleavable sites in the primer. Such cleavable sites also eliminate unnecessary sequences and allow for the use of a capture and release assay and for primer modification for the interleaving multiplexing method, described herein. For example, moving the cleavable site along the primer backbone directly modifies the mass of the PCR™ product. The cleavable site is

typically introduced via a cleavable primer and the cleavable site is located outside of the region of interest. Cleavable primers may include those comprising an exonuclease blocking moiety, a Type IIS restriction endonuclease recognition site, and a Type II restriction endonuclease recognition site.

5       The target nucleic acids may, thus, be reduced in length by any of the methods known that will cleave within one or more flanking regions preferably without cleaving within the region of interest. Exemplary methods of reducing length include: cleaving at endogenous restriction endonuclease cleavable sites present in one or more flanking regions but absent in the region of interest; cleaving at restriction  
10       endonuclease cleavable sites at or adjacent to restriction endonuclease recognition sites incorporated into one or more flanking regions by use of one or more cleavable primers comprising said restriction endonuclease recognition sites; cleaving at a combination of restriction endonuclease cleavable sites wherein the sites are endogenous and/or introduced using mismatch or overhanging primers; and selective  
15       digestion of one or more flanking regions using exonuclease and an exonuclease blocking moiety to protect the regions of interest from digestion.

      The restriction endonucleases employed with the present invention include type II and type IIS restriction endonucleases. The restriction endonuclease recognition sites may be either within a primer region, or outside the primer region, so  
20       long as the restriction endonuclease cleavable sites are within one or more flanking regions and preferably not within a region of interest. For type II restriction endonucleases, the restriction endonuclease recognition site is the same as the restriction endonuclease cleavable site. For Type IIS restriction endonucleases, the cleavable site is at a defined distance away from one side of the recognition site.

25       Another embodiment of the invention involves using a cleavable primer having an exonuclease blocking moiety. After amplification of the target nucleic acid, the amplified target nucleic acid will include an exonuclease blocking moiety. The amplified target nucleic acid is then treated with a 5' to 3' exonuclease, which degrades the strand containing the exonuclease blocking moiety in a 5' to 3' direction  
30       only up to the blocking moiety. The 5' to 3' exonuclease may optionally degrade the other complementary strand of the amplified target nucleic acid, in cases where the



other strand does not have an exonuclease blocking moiety. The treatment with the 5' to 3' exonuclease leaves a reduced-length, single-stranded amplified target nucleic acid for mass spectrometric analysis.

Cleavable sites may also include chemically cleavable groups incorporated within the phosphate backbone linkage (e.g. replacement of phosphate with a phosphoramidate) or as a substituent on or replacement of one of the bases or sugars of the oligonucleotide primer (e.g. a modified base or sugar, for example, a more labile glycosidic linkage). Such chemically cleavable groups would be apparent to one of skill in the art in light of the present disclosure and include, for example, dialkoxysilane, 3'-(S)-phosphorothioate, 5'-(S)-phosphorothioate, 3'-(N)-phosphoroamidate, 5'-(N)-phosphoroamidate, and ribose. FIGS. 16A and 16B depict a 3'-(S)-phosphorothioate and 5'-(S)-phosphorothioate, respectively as defined in this invention. Note that these linkages are often referred to as thiophosphates as well. A similar nomenclature is employed for 3'-(N)-phosphoroamidate, 5'-(N)-phosphoroamidate. The chemically cleavable site should generally be stable under the amplification, hybridization and washing conditions to be employed and is preferably within one or more of the flanking regions.

In a preferred embodiment, the cleavable site is located near the 3' end of the primer used to bind the amplified target nucleic acid to the solid support. By locating the cleavable site near the 3' end, it is possible to further reduce the length of the amplified target nucleic acid, eliminating a flanking region from the polynucleotide region of interest. Cleavable primers are described in PCT/US96/06116, filed April 26, 1996 (incorporated herein by reference).

The primer pairs described in this invention may be combined to generate overlapping PCR™ product sizes which are all distinguishable by mass.

### EXAMPLE 1

#### PCR CONDITIONS FOR MULTIPLEXING DTNR RESULTS

**Template:** 5  $\mu$ l, 1:1000 dilution of AmpFISTR Green I Allelic Ladders (PE Applied Biosystems; contains common alleles from the STR loci CSF1PO, TPOX,

and TH01 and the sex-typing marker amelogenin); for regular samples, 2-5  $\mu$ L of 1-10 ng of human genomic DNA was added to the PCR reaction.

**Reaction Mix:** 20  $\mu$ L reaction with 1X STR buffer (Promega; contains 1.5 mM  $MgCl_2$ , 200  $\mu$ M dNTPs, etc.), 1 U Taq polymerase (Promega), 20 pmol forward  
5 and reverse primers with one of them being a primer containing a biotin moiety on the 5'-end and a thiothymine residue near the 3'-end of the oligonucleotide.

**Thermal Cycling:** In 0.2 mL tubes in an MJ Research DNA Engine (block temperature) 94 °C for 2 min; 35 cycles: 94 °C for 30 sec, 60 °C for 30 sec, 72 °C for 30 sec; 72 °C for 5 min.

10

## EXAMPLE 2

### SAMPLE PURIFICATION FOR MULTIPLEXING DTNR RESULTS

A typical binding/washing protocol for purifying samples for DTNR  
15 multiplexing includes the following steps:

- a) Wash 10  $\mu$ L streptavidin-coated magnetic beads with 2X binding/wash buffer
- b) Repeat a second time
- c) Add 5  $\mu$ L 5X binding/wash buffer then add ~19  $\mu$ L of PCR sample to the  
20 beads (1  $\mu$ L was removed for an agarose gel check) and vortex sample tube for 15 min at slow speed
- d) Wash beads with 30  $\mu$ L of 2X binding/wash buffer
- e) Wash beads with 30  $\mu$ L of 0.1 N NaOH
- f) Add 30  $\mu$ L of 0.1 N NaOH and vortex for 10 min at slow speed
- 25 g) Wash beads with 30  $\mu$ L of 0.1 N NaOH
- h) Wash beads with 30  $\mu$ L of 20 mM ammonium acetate
- i) Repeat step (h) five times
- j) Wash beads with deionized water
- k) Repeat step (j) twice

30 Note after each step, the supernatant is removed while the beads are magnetically held in the bottom of the tube.

After purification the solid bound strands were released by cleaving at the chemically cleavable thiophosphate site by the following procedure: 7  $\mu$ L of 0.1 mM silver nitrate was added and the samples were incubated at 48 °C for 15 min.; the supernatant was then transferred to a clean tube and 2  $\mu$ L of 70 mM DTT was added; and finally the sample was dried in a speed vacuum. For mixed samples the preceding protocol was modified in that aliquots of the samples (e.g., 3  $\mu$ L TII01 ladder + 3  $\mu$ L TPOX ladder) were mixed before the drying step.

### EXAMPLE 3

#### 10 MS ANALYSIS FOR MULTIPLEXING DTNR RESULTS

The matrix consisted of a 5:1 molar ratio of 3-hydroxypicolinic acid (3-HPA; Lancaster Synthesis) to picolinic acid (PA; Aldrich) and was prepared by mixing 18  $\mu$ L of a freshly prepared saturated 3-HPA solution (~0.5 M) with 2  $\mu$ L of 1 M PA.

The sample to be analyzed was reconstituted in 0.5  $\mu$ L of the matrix and manually spotted on the sample plate.

The instrument conditions employed with a linear time-of-flight mass spectrometer consisted of the following: acceleration voltage of +20 kV; delay of +3.6 kV at 1.12  $\mu$ sec; laser setting of 179 on the polarizer; mass gate of 5.84  $\mu$ sec; and 400 shots. A 2-point mass calibration with a 15-mer (4507.0 Da) and a 36-mer (10998.2 Da) was employed.

### EXAMPLE 4

Oligonucleotide primers are typically prepared by the phosphoramidite approach. In this automated, solid-phase procedure, each nucleotide is individually added to the 5'-end of the growing oligonucleotide chain, which is in turn attached at the 3'-end to a solid support. The added nucleotides are in the form of trivalent 3'-phosphoramidites that are protected from polymerization by a dimethoxytrityl ("DMT") group at the 5'-position. After base induced phosphoramidite coupling, mild oxidation to give a pentavalent phosphotriester intermediate and DMT removal provides a new site for oligonucleotide elongation. These syntheses may be performed on a Perkin Elmer/Applied Biosystems Division DNA synthesizer. The

oligonucleotide primers are then cleaved off the solid support, and the phosphodiester and exocyclic amino groups are deprotected with ammonium hydroxide.

The biotin, and 3'- and 5'-(S) phosphorothioate linkages are also prepared in an automated fashion from phosphoramidite intermediates using similar procedures  
5 and either modified bases or activated and protected linker molecules.

### EXAMPLE 5

#### TWO STAGE MULTIPLEXING REACTION:

#### OUTSIDE PRIMERS FOR HIGHER LEVEL MULTIPLEX FOLLOWED BY 10 SINGLE DDN TERMINATION

A triplex PCR reaction was run with 10-ng human genomic DNA template in a 20- $\mu$ L PCR reaction. Primers specific for the three STR loci CSF1PO, TPOX, and TH01 were used as described by Huang *et al*. These primers produce larger sized  
15 PCR products than the primers listed in this patent and the primer sequences from Table 1 for these three STR loci are within the product region.

Multiplex PCR components: 20  $\mu$ L reaction containing 1.5X Taq buffer II (PE Applied Biosystems), 200  $\mu$ M dNTPs, 1.5 mM  $MgCl_2$ , 1 U AmpliTaq Gold (PE Applied Biosystems), 0.5  $\mu$ M each primer

20 Thermal cycling was performed in 0.2 mL tubes using an MJ Research DNA Engine (calculated temperature) with the following cycling parameters: 95 °C for 11 min; 40 cycles: 94 °C for 30 sec, 64 °C for 30 sec, 68 °C for 45 sec; 70 °C for 10 min.

Following PCR, the sample was treated with shrimp-alkaline phosphatase (SAP) to hydrolyze the unincorporated dNTPs. Typically, 1 U SAP was added to  
25 each 20  $\mu$ L PCR sample. The sample was then incubated at 37 °C for 60 minutes followed by heating at 75 °C for 15 minutes.

A multiplexed primer extension assay was then performed using cleavable primers for the three STR loci. The reaction included three regular deoxynucleotide triphosphates (dNTP) to allow incorporation through the STR repeat region and a  
30 single dideoxynucleotide triphosphate (ddNTP) to halt extension (see Braun, et al.). The termination by the ddNTP shortened the measured STR allele.

Multiplexed primer extension reaction components: 20  $\mu$ l. reaction containing 1X TaqI'S buffer (PE Applied Biosystems), 2.4 U TaqFS, 200  $\mu$ M dGTP, 200  $\mu$ M dATP, 200  $\mu$ M dTTP, 400  $\mu$ M ddCTP, 40 pmol CSF1PO primer (SEQ ID NO:100), 20 pmol TPOX (SEQ ID NO:31), 20 pmol TH01 (SEQ ID NO:103), and 2  $\mu$ l. SAP-treated PCR product (as described above).

Thermal cycling for extension reaction was performed in 0.2 mL tubes using an MJ Research DNA Engine (calculated temperature) with the following cycling parameters: 95 °C for 2 min; 30 cycles: 94 °C for 30 sec, 55 °C for 30 sec, 72 °C for 45 sec; 60 °C for 5 min. The resultant product was purified and analyzed as detailed in the examples above.

As seen in Figure 8A, the possible alleles including microvariants, such as TH01 9.3, are resolvable from one another in all three STR systems even though they overlap in the mass range. Figure 8B illustrates a result with this particular STR multiplex. The sample contains a homozygous TPOX allele 8, heterozygous TH01 alleles 6 and 9.3, and a homozygous CSF1PO allele 12. In this particular case, the strand containing an AATG or ATAG repeat was used in all three STR loci so that the same dideoxynucleotide triphosphate (ddC) could be used to terminate the strand on the opposite side of the repeat from the cleavable primer. After the extension reaction with the ddNTP and the cleavage reaction, the repeat region and only a flanking few bases on either side of the repeat are present for the three STR loci. Mass accuracy as well as resolution and sensitivity are improved in the mass spectrometer by going to smaller sizes for the STR alleles using this approach.

#### **EXAMPLE 6**

**TWO STAGE MULTIPLEXING REACTION:**  
**OUTSIDE PRIMERS FOR HIGHER LEVEL MULTIPLEX FOLLOWED BY**  
**GTS PRIMERS IN LOWER LEVEL MULTIPLEX THAT PRODUCE**  
**SMALLER PCR PRODUCTS**

In a situation where there is a small or limited amount of DNA template available, a highly multiplexed PCR reaction may be performed initially followed by

a second stage PCR with primers that are interior (i.e., closer to the repeat region) than those contained in the first multiplex ("nested PCR"). The first multiplex could include, for example, 8-14 STR loci that are PCR-amplified simultaneously. Aliquots of these amplicons could then be divided and reamplified in a second PCR reaction with primers for a subset of the STR loci already amplified. For example, six duplex reactions or four triplexes with mass spectroscopy primers could be performed to measure all of the STR loci in an initial 12-plex reaction. Primers used in these duplexes could be from those listed in Table 1. Provided that the primers for the first stage multiplex are outside (or at least identical to) the second stage primer sets, this approach will work for any PCR-compatible primers.

The advantage of the nested PCR approach is that a SAP-treatment is not required (as in Example 5) before the second stage reaction as dNTPs are used. However, measured STR alleles will be longer and thus more challenging to analyze in the mass spectrometer than STR allele products created with the primer termination reaction (ddN) approach described above.

#### **EXAMPLE 7**

##### **FTA PAPER USED IN PCR REACTIONS IN PLACE OF EXTRACTED DNA**

Liquid blood was stained on an FTA<sup>TM</sup> Card (Life Technologies, Gaithersburg, MD) and allowed to air-dry for 1 hour. A small portion of blood-stained paper was cut out of the center of the spot and placed in a 0.6 mL tube. An aliquot of 50  $\mu$ L FTA<sup>TM</sup> Purification Reagent (Life Technologies) was added to the tube and vortexed for several seconds. The tube was allowed to sit for 5 minutes at room temperature. The tube was vortexed again and the liquid in the tube was removed. Another aliquot of 50  $\mu$ L FTA<sup>TM</sup> Purification Reagent was added to the tube and the vortexing and waiting steps were repeated as described above. The FTA<sup>TM</sup> paper was then washed a third time with FTA<sup>TM</sup> Purification Reagent and then washed twice more with deionized water. After these washes, the liquid was removed with a pipet and the FTA<sup>TM</sup> paper punch was left in the tube to air-dry for 1 hour.

- The dried punch was added directly to PCR amplification mix in place of extracted human genomic DNA. PCR was performed as described in the other examples with no changes to amplification conditions or reagents. No decrease in PCR efficiency was observed when the FTA<sup>TM</sup> paper was compared to standard K562
- 5 human genomic DNA templates. The use of FTA<sup>TM</sup> paper greatly facilitates the automation of DNA typing procedures as extensive DNA extraction is not needed from liquid blood samples.

### REFERENCES

The following references, to the extent that they provide exemplary procedural or other details supplementary to those set forth herein, are specifically incorporated herein by reference.

5

U.S. Patent No. 4,683,202 Mullis

U.S. Patent No. 5,364,759 Caskey et al.

U.S. Patent No. 5,378,602 Polymeropoulos et al.

U.S. Patent No. 5,599,666 Schumm et al.

10 

U.S. Patent No. 5,605,798 Köster

U.S. Patent No. 5,700,642 Monforte et al.

U.S. Patent No. 5,674,686 Schumm and Pucrs

U.S. Patent No. 5,766,847 Jackle and Tautz

U.S. Patent No. 5,496,562 Burgoyne

15

Alford, Hammond, Coto, Caskey, "Rapid and efficient resolution of parentage by amplification of short tandem repeats," *Am. J. Hum. Genet.*, 55: 190-195, 1994.

20 

Anker, Steinbreck, Donis-Keller, "Tetranucleotide repeat polymorphism at the human thyroid peroxidase (hTPO) locus," *Hum. Mol. Genet.*, 1:137, 1992.

Becker, Li, Shaler, Hunter, Lin, Monforte, "Genetic analysis of short tandem repeat loci by time of flight mass spectrometry," *Seventh International Symposium on Human Identification (1996)*, pp. 158-162, 1997.

25 

Dubovsky, Sheffield, Dzyk, Weber, "Sets of short tandem repeat polymorphisms for efficient linkage screening of the human genome," *Hum. Mol. Genet.*, 4: 449-452, 1995.

Edwards, Civitello, Hammond, Caskey, "DNA typing and genetic mapping with trimeric and tetrameric tandem repeats," *Am. J. Hum. Genet.*, 49:746-756, 1991.



- Fregeau and Fourney, "DNA typing with fluorescently tagged short tandem repeats: a sensitive and accurate approach to human identification," *BioTechniques*, 15:100-119, 1993.
- Hammond, Jin, Zhong, Cuskey, Chakraborty, "Evaluation of 13 short tandem repeat loci for use in personal identification applications," *Am. J. Hum. Genet.*, 55:175-189, 1994.
- Hauge and Liu, "A study of the origin of 'shadow bands' seen when typing dinucleotide repeat polymorphisms by the PCR™," *Hum. Mol. Genet.*, 2:411-415, 1993.
- 10 Hearne and Todd, "Tetranucleotide repeat polymorphism at the HPRT locus," *Nucleic Acids Res.*, 19:5450, 1991.
- Kimpton, Walton, Gill, "A further tetranucleotide repeat polymorphism in the vWF gene," *Hum. Mol. Genet.*, 1:287, 1992.
- Kimpton, Gill, Walton, Urquhart, Millican, Adams, "Automated DNA profiling employing multiplex amplification of short tandem repeat loci," *PCR™ Meth. Appl.*, 3:13-22, 1993.
- 15 Kimpton, Oldroyd, Watson, Frazier, Johnson, Millican, Urquhart, Sparkes, Gill, "Validation of highly discriminating multiplex short tandem repeat amplification systems for individual identification," *Electrophoresis*, 17:1283-1293, 1996.
- 20 Lareu, Pestoni, Schurenkamp, Rand, Brinkmann, Carracedo, "A highly variable STR at the D12S391 locus," *Int. J. Leg. Med.*, 109:134-138, 1996.
- Lygo, Johnson, Holdaway, Woodroffe, Whitaker, Clayton, Kimpton, Gill, "The validation of short tandem repeat (STR) loci for use in forensic casework," *Int. J. Leg. Med.*, 107:77-89, 1994.
- 25 Polymenopoulos, Rath, Xiao, Merrill, "Tetranucleotide repeat polymorphism at the human c-fes/fps proto-oncogene (FES)," *Nucleic Acids Res.*, 19:4018, 1991.
- Polymenopoulos, Rath, Xiao, Merrill, "Tetranucleotide repeat polymorphism at the human coagulation factor XIII A subunit gene (F13A1)," *Nucleic Acids Res.*, 19:4306, 1991.
- 30

- Polymeropoulos, Xiao, Rath, Merril, "Tetranucleotide repeat polymorphism at the human tyrosine hydroxylase gene (TH)," *Nucleic Acids Res.*, 19:3753, 1991.
- Puers, Hammond, Caskey, Lins, Sprecher, Brinkmann, Schumm, "Allele ladder characterization of the short tandem repeat polymorphism located in the 5' flanking region to the human coagulation factor XIII A subunit gene," *Genomics*, 23:260-264, 1994.
- Puers, Hammond, Jin, Caskey, Schumm, "Identification of repeat sequence heterogeneity at the polymorphic short tandem repeat locus HUM11101[AATG]<sub>n</sub> and reassignment of alleles in population analysis by using a locus-specific allele ladder," *Am. J. Hum. Genet.*, 53:953-958, 1993.
- Rucwicz, Arnenmann, Spurr, Grzeschik, Epplen, "Simple repeat sequences on the human Y chromosome are equally polymorphic as their autosomal counterparts," *Hum. Genet.*, 89:389-394, 1992.
- The Utah Marker Development Group "A collection of ordered tetranucleotide-repeat markers from the human genome," *Am. J. Hum. Genet.*, 57:619-628, 1995.
- Weber and May, "Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction," *Am. J. Hum. Genet.*, 44:388-396, 1989.
- Ziegler, Su, Corcoran, Nie, Mayrand, Hoff, McBride, Kronick, Diehl, "Application of automated DNA sizing technology for genotyping microsatellite loci," *Genomics*, 14:1026-1031, 1992.
- Braun, A., et al., "Detecting CFTR gene mutations by using primer oligo base extension and mass spectrometry," *Clin. Chem.*, 43:1151-1158 (1997).
- Braun, A., et al., "Improved Analysis of Microsatellites Using Mass Spectrometry," *Genomics* 46:18-23 (1997).
- Butler, J.M., et al., "Reliable Genotyping of Short Tandem Repeat Loci without an Allelic Ladder Using Time-of-Flight Mass Spectrometry," *Int. J. Legal Med.*, in press (1998).
- Butler, J.M., et al., "Rapid and Automated Analysis of Short Tandem Repeat Loci Using Time-of-Flight Mass Spectrometry," *Proceedings of the Eighth International Symposium on Human Identification 1997*, Promega Corporation, pp. 94-101 (1998).

- Butler, J.M., et al., "High-throughput STR Analysis by Time-of-Flight Mass Spectrometry," *Proceedings of the Second European Symposium on Human Identification 1998*, Promega Corporation, in press (1998).
- Huang, N.F., et al., "Chinese population data on three tetrameric short tandem repeat loci—TH01, TPOX, and CSF1PO—derived using multiplex PCR and manual typing," *Forensic Sci. Int.* 71:131-136 (1995).
- Kayser, M., et al., "Evaluation of Y-chromosomal STRs: a multicenter study," *Int. J. Legal Med.* 110:125-133 (1997).
- Little, D.P., et al., "MALDI on a Chip: Analysis of Arrays of Low-Femtomole to Subfemtomole Quantities of Synthetic Oligonucleotides and DNA Diagnostic Products Dispensed by a Piezoelectric Pipet," *Anal. Chem.* 69:4540-4546 (1997).
- Little, D.P., et al., "Mass Spectrometry from Miniaturized Arrays for Full Comparative DNA Analysis," *Nature Med.* 3:1413-1416 (1997).
- Ross, P.L., and Belgrader, P., "Analysis of Short Tandem Repeat Polymorphisms in Human DNA by Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry," *Anal. Chem.* 69:3966-3972 (1997).
- Ross, P.L., et al., "Analysis of DNA Fragments from Conventional and Microfabricated PCR Devices Using Delayed Extraction MALDI-TOF Mass Spectrometry," *Anal. Chem.* 70:2067-2073 (1998).
- Taranenko, N.I., et al., "Matrix-assisted Laser Desorption/Ionization for Short Tandem Repeat Loci," *Rapid Commun. Mass Spectrom.* 12:413-418 (1998).
- Wenz, H.-M., et al., "High-Precision Genotyping by Denaturing Capillary Electrophoresis," *Genome Res.* 8:69-80 (1998).

**WHAT IS CLAIMED IS:**

1. A method for analyzing DNA tandem nucleotide repeat alleles at a DNA tandem nucleotide repeat locus in a target nucleic acid by mass spectrometry, said method comprising
  - a) obtaining a target nucleic acid comprising a DNA tandem nucleotide repeat region;
  - b) extending the target nucleic acid using one or more primers to obtain a limited size range of nucleic acid extension products, wherein one or more primers are complementary to a sequence flanking the DNA tandem nucleotide repeat of said locus; and
  - c) determining the mass of the nucleic acid extension products by mass spectrometry.
2. The method of claim 1, wherein a 3' end of one or more primers immediately flanks a DNA tandem nucleotide repeat region.
3. The method of claim 1, wherein one or more primers comprise a sequence complementary to up to one tandem repeat of the DNA tandem nucleotide repeat locus.
4. The method of claim 3, wherein one or more primers comprise a sequence complementary to up to two tandem repeats of the DNA tandem nucleotide repeat locus.
5. The method of claim 4, wherein one or more primers comprise a sequence complementary to up to three tandem repeats of the DNA tandem nucleotide repeat locus.
6. The method of claim 1, wherein at least one primer comprises a cleavable site.

7. The method of claim 6, wherein the cleavable site comprises a recognition site for a restriction endonuclease, an exonuclease blocking site, or a chemically cleavable site.
8. The method of claim 7, wherein the cleavable site comprises a chemically cleavable site.
- 5 9. The method of claim 6, wherein wherein at least one primer is capable of attaching to a solid support.
- 10 10. The method of claim 9, wherein at least one primer comprises biotin or digoxigenin.
11. The method of claim 1, wherein the extension of at least one primer is terminated using a chain termination reagent.
12. The method of claim 12, wherein the chain termination reagent is a dideoxynucleotide triphosphate.
- 15 13. The method of of claim 1, wherein the locus is selected from the group consisting of CSF1PO, D5S818, D7S820, D8S1179, D13S327, D16S539, D16S539, D18S51, D21S11, DYS19, F13A1, FES/FPS, FGA, HPR1B, TH01, TPOX, DYS388, DYS391, DYS392, DYS393, D2S1391, D18S535, D2S1338, D19S433, D6S477, 20 D1S518, D14S306, D22S684, F13B, CD4, D12S391, D10S220 and D7S523.
14. The method of claim 13, wherein two of the primers comprise a pair of sequences selected from the group consisting of
- a sequence according to SEQ ID NO:1 and a sequence according to SEQ ID NO:2;
  - a sequence according to SEQ ID NO:3 and a sequence according to SEQ ID NO:4;

a sequence according to SEQ ID NO:5 and a sequence according to SEQ ID NO:6;

a sequence according to SEQ ID NO:7 and a sequence according to SEQ ID NO:8;

a sequence according to SEQ ID NO:9 and a sequence according to SEQ ID NO:10;

a sequence according to SEQ ID NO:11 and a sequence according to SEQ ID NO:12;

a sequence according to SEQ ID NO:13 and a sequence according to SEQ ID NO:14;

a sequence according to SEQ ID NO:15 and a sequence according to SEQ ID NO:16;

a sequence according to SEQ ID NO:17 and a sequence according to SEQ ID NO:18;

a sequence according to SEQ ID NO:19 and a sequence according to SEQ ID NO:20;

a sequence according to SEQ ID NO:21 and a sequence according to SEQ ID NO:22;

a sequence according to SEQ ID NO:23 and a sequence according to SEQ ID NO:24;

a sequence according to SEQ ID NO:25 and a sequence according to SEQ ID NO:26;

a sequence according to SEQ ID NO:27 and a sequence according to SEQ ID NO:28;

a sequence according to SEQ ID NO:29 and a sequence according to SEQ ID NO:30;

a sequence according to SEQ ID NO:31 and a sequence according to SEQ ID NO:32;

a sequence according to SEQ ID NO:49 and a sequence according to SEQ ID NO:83;

a sequence according to SEQ ID NO:52 and a sequence according to SEQ ID NO:84;

a sequence according to SEQ ID NO:54 and a sequence according to SEQ ID NO:85;

a sequence according to SEQ ID NO:56 and a sequence according to SEQ ID NO:86;

a sequence according to SEQ ID NO:58 and a sequence according to SEQ ID NO:87;

a sequence according to SEQ ID NO:59 and a sequence according to SEQ ID NO:88;

a sequence according to SEQ ID NO:62 and a sequence according to SEQ ID NO:89;

a sequence according to SEQ ID NO:63 and a sequence according to SEQ ID NO:90;

a sequence according to SEQ ID NO:66 and a sequence according to SEQ ID NO:91;

a sequence according to SEQ ID NO:67 and a sequence according to SEQ ID NO:92;

a sequence according to SEQ ID NO:70 and a sequence according to SEQ ID NO:93;

a sequence according to SEQ ID NO:72 and a sequence according to SEQ ID NO:94;

a sequence according to SEQ ID NO:74 and a sequence according to SEQ ID NO:95;

a sequence according to SEQ ID NO:76 and a sequence according to SEQ ID NO:96;

a sequence according to SEQ ID NO:78 and a sequence according to SEQ ID NO:97;

a sequence according to SEQ ID NO:80 and a sequence according to SEQ ID NO:98;

a sequence according to SEQ ID NO:66 and a sequence according to SEQ ID NO:99;

a sequence according to SEQ ID NO:33 and a sequence according to SEQ ID NO:100 and

a sequence according to SEQ ID NO:101 and a sequence according to SEQ ID NO:103.

15. The method of claim 14, wherein more than one target nucleic acid is extended to produce more than one nucleic acid extension product.

16. The method of claim 15, wherein the mass of more than one DNA tandem nucleotide repeat allele at more than one DNA tandem nucleotide repeat loci are determined simultaneously.

17. The method of claim 16, wherein the DNA tandem nucleotide repeat loci have overlapping allelic mass ranges.

18. The method of claim 17, wherein the nucleic acid extension products have interleaving mass spectral peaks.

19. The method of claim 17, wherein at least one nucleic acid extension product contains a mass modified nucleotide.

20. The method of claim 17, wherein the length of at least one nucleic acid extension product is reduced by cleaving the nucleic acid extension product at a cleavable site.

21. The method of claim 20, wherein the cleavable site comprises a restriction endonuclease cleavage site, an exonuclease blocking site, or a chemically cleavable group.



22. The method of claim 21, wherein the cleavable site comprises a recognition site for a restriction endonuclease.
23. The method of claim 21, wherein the cleavable site comprises an exonuclease blocking site.
24. The method of claim 21, wherein the cleavable site comprises a chemically cleavable site.
25. The method of claim 20, wherein the first primer is capable of attaching to a solid support.
26. The method of claim 25, wherein the first primer comprises biotin or digoxigenin.
27. A method for multiplexing the identification of more than one DNA tandem nucleotide repeat region from more than one DNA tandem nucleotide repeat loci, comprising:
- obtaining more than one nucleic acid extension product by extending one or more primers complementary to sequences flanking the DNA tandem nucleotide repeat region, and;
- determining the masses of more than one nucleic acid extension product simultaneously by mass spectrometry,
- wherein the nucleic acid extension products have overlapping allelic mass ranges.
28. The method of claim 27, wherein a 3' end of one or more primers immediately flanks a DNA tandem nucleotide repeat region.
29. The method of claim 27, wherein one or more primers comprise a sequence complementary to up to one tandem repeat of the DNA tandem nucleotide repeat locus.

30. The method of claim 29, wherein one or more primers comprise a sequence complementary to up to two tandem repeats of the DNA tandem nucleotide repeat locus.
- 5 31. The method of claim 30, wherein one or more primers comprise a sequence complementary to up to three tandem repeats of the DNA tandem nucleotide repeat locus.
- 10 32. The method of claim 27, wherein the extension of at least one primer is terminated using a chain termination reagent.
33. The method of claim 32, wherein the chain termination reagent is a dideoxynucleotide triphosphate.
- 15 34. The method of claim 32, wherein at least one target nucleic acid extension product contains a mass modifying group.
35. The method of claim 34, wherein the mass modifying group comprises a mass modified nucleotide.
- 20 36. The method of claim 34, wherein the mass modifying group comprises a nonstandard deoxyribonucleotide.
37. The method of claim 27, wherein the cleavable site comprises a recognition site for a restriction endonuclease, an exonuclease blocking site, or a chemically cleavable site.
- 25 38. The method of claim 37, wherein the cleavable site comprises a recognition site for a restriction endonuclease.

39. The method of claim 37, wherein the cleavable site comprises an exonuclease blocking site.
40. The method of claim 37, wherein the cleavable site comprises a chemically  
5 cleavable site.
41. The method of claim 34, wherein the mass modifying group is incorporated during extension of the nucleic acid extension product.
42. The method of claim 34, wherein the mass modifying group is incorporated  
10 after extension of the nucleic acid extension product.
43. A method for multiplexing the identification of more than one DNA tandem nucleotide repeat region from more than one DNA tandem nucleotide repeat loci, comprising:
- 15 obtaining more than one nucleic acid amplification product by amplifying two or more primers complementary to sequences flanking the DNA tandem nucleotide repeat region; and  
determining the masses of more than one nucleic acid amplification product simultaneously by mass spectrometry,  
20 wherein the nucleic acid extension products have overlapping allelic mass ranges.
44. The method of claim 27, wherein a 3' end of one or more primers immediately flanks a DNA tandem nucleotide repeat region.
- 25 45. The method of claim 27, wherein one or more primers comprise a sequence complementary to up to one tandem repeat of the DNA tandem nucleotide repeat locus.
- 30 46. The method of claim 45, wherein one or more primers comprise a sequence complementary to up to two tandem repeats of the DNA tandem nucleotide repeat locus.

47. The method of claim 46, wherein one or more primers comprise a sequence complementary to up to three tandem repeats of the DNA tandem nucleotide repeat locus.

5

48. The method of claim 27, wherein at least one target nucleic acid amplification product contains a mass modifying group.

49. The method of claim 34, wherein the mass modifying group comprises a mass modified nucleotide.

10

50. The method of claim 34, wherein the mass modifying group comprises a nonstandard deoxyribonucleotide.

51. The method of claim 27, wherein the cleavable site comprises a recognition site for a restriction endonuclease, an exonuclease blocking site, or a chemically cleavable site.

15

52. The method of claim 37, wherein the cleavable site comprises a recognition site for a restriction endonuclease.

53. The method of claim 37, wherein the cleavable site comprises an exonuclease blocking site.

20

54. The method of claim 37, wherein the cleavable site comprises a chemically cleavable site.

55. The method of claim 34, wherein the mass modifying group is incorporated during amplification.

25

56. The method of claim 34, wherein the mass modifying group is incorporated after amplification.
57. A kit for analyzing alleles of a DNA tandem nucleotide repeat locus in a target  
5 nucleic acid by mass spectrometry, comprising  
a first primer complementary to a sequence flanking a DNA tandem nucleotide  
repeat region of said locus, and  
a second primer complementary to a sequence flanking a DNA tandem  
nucleotide repeat region of said locus,  
10 wherein the first primer and second primer are used to generate a limited size range of  
nucleic acid products for analysis by mass spectrometry.
58. The method of claim 57, wherein a 3' end of one or more primers immediately  
flanks a DNA tandem nucleotide repeat region.
- 15 59. The method of claim 57, wherein a 3' end of one or more primers immediately  
flank a DNA tandem nucleotide repeat region.
60. The kit of claim 57, wherein at least one of the first primer or the second  
primer comprise a sequence complementary to up to one tandem repeat of the DNA  
20 tandem nucleotide repeat locus.
61. The kit of claim 57, wherein at least one of the first primer or the second  
primer comprise a sequence complementary to up to two tandem repeats of the DNA  
tandem nucleotide repeat locus.
- 25 62. The kit of claim 57, wherein at least one of the first primer or the second  
primer comprise a sequence complementary to up to three tandem repeats of the DNA  
tandem nucleotide repeat locus.
63. The kit of claim 57, wherein the first primer comprises a cleavable site.

64. The kit of claim 63, wherein the cleavable site comprises a recognition site for a restriction endonuclease, an exonuclease blocking site, or a chemically cleavable site.

65. The kit of claim 64, wherein the cleavable site comprises a recognition site for a restriction endonuclease.

5

66. The kit of claim 64, wherein the cleavable site comprises a chemically cleavable site.

67. The kit of claim 63, wherein wherein the first primer is capable of attaching to a solid support.

10

68. The kit of claim 67, wherein the first primer comprises biotin or digoxigenin.

69. The kit of claim 57 wherein the DNA tandem nucleotide repeat locus is selected from the group consisting of CSF1PO, D5S818, D7S820, D8S1179, D13S317, D16S539, D16S539, D18S51, D21S11, DYS19, F13A1, FES/FPS, FGA, HPRTB, TH01, TPOX, DYS388, DYS391, DYS392, DYS393, D2S1391, D18S535, D2S1338, D19S433, D6S477, D1S518, D14S306, D22S684, F13B, CD4, D12S391, D10S220 and D7S523.

70. The kit of claim 69, wherein the first and second primer comprise a pair of sequences selected from the group consisting of

a sequence according to SEQ ID NO:1 and a sequence according to SEQ ID NO:2;

a sequence according to SEQ ID NO:3 and a sequence according to SEQ ID NO:4;

a sequence according to SEQ ID NO:5 and a sequence according to SEQ ID NO:6;

a sequence according to SEQ ID NO:7 and a sequence according to SEQ ID NO:8;

a sequence according to SEQ ID NO:9 and a sequence according to SEQ ID NO:10;

a sequence according to SEQ ID NO:11 and a sequence according to SEQ ID NO:12;

a sequence according to SEQ ID NO:13 and a sequence according to SEQ ID NO:14;

a sequence according to SEQ ID NO:15 and a sequence according to SEQ ID NO:16;

a sequence according to SEQ ID NO:17 and a sequence according to SEQ ID NO:18;

a sequence according to SEQ ID NO:19 and a sequence according to SEQ ID NO:20;

a sequence according to SEQ ID NO:21 and a sequence according to SEQ ID NO:22;

a sequence according to SEQ ID NO:23 and a sequence according to SEQ ID NO:24;

a sequence according to SEQ ID NO:25 and a sequence according to SEQ ID NO:26;

a sequence according to SEQ ID NO:27 and a sequence according to SEQ ID NO:28;

a sequence according to SEQ ID NO:29 and a sequence according to SEQ ID NO:30;

a sequence according to SEQ ID NO:31 and a sequence according to SEQ ID NO:32;

a sequence according to SEQ ID NO:49 and a sequence according to SEQ ID NO:83;

a sequence according to SEQ ID NO:52 and a sequence according to SEQ ID NO:84;

a sequence according to SEQ ID NO:54 and a sequence according to SEQ ID NO:85;

a sequence according to SEQ ID NO:56 and a sequence according to SEQ ID NO:86;

a sequence according to SEQ ID NO:58 and a sequence according to SEQ ID NO:87;

a sequence according to SEQ ID NO:59 and a sequence according to SEQ ID NO:88;

a sequence according to SEQ ID NO:62 and a sequence according to SEQ ID NO:89;

a sequence according to SEQ ID NO:63 and a sequence according to SEQ ID NO:90;

a sequence according to SEQ ID NO:66 and a sequence according to SEQ ID NO:91;

a sequence according to SEQ ID NO:67 and a sequence according to SEQ ID NO:92;

a sequence according to SEQ ID NO:70 and a sequence according to SEQ ID NO:93;

a sequence according to SEQ ID NO:72 and a sequence according to SEQ ID NO:94;

a sequence according to SEQ ID NO:74 and a sequence according to SEQ ID NO:95;

a sequence according to SEQ ID NO:76 and a sequence according to SEQ ID NO:96;

a sequence according to SEQ ID NO:78 and a sequence according to SEQ ID NO:97;

a sequence according to SEQ ID NO:80 and a sequence according to SEQ ID NO:98;



a sequence according to SEQ ID NO:66 and a sequence according to SEQ ID NO:99;

a sequence according to SEQ ID NO:53 and a sequence according to SEQ ID NO:100 and

a sequence according to SEQ ID NO:101 and a sequence according to SEQ ID NO:103.

71. A kit for analyzing alleles of more than one DNA tandem nucleotide repeat locus in a target nucleic acid by mass spectrometry, comprising  
a plurality of primers, wherein each said primer is complementary to a  
5 sequence flanking a DNA tandem nucleotide repeat region and selected  
to generate a limited size range of nucleic acid products for analysis of  
said locus by mass spectrometry upon extension.
72. The method of claim 71, wherein a 3' end of one or more primers immediately  
10 flanks a DNA tandem nucleotide repeat region.
73. The kit of claim 71, wherein at least one of the plurality of primers comprises  
a sequence complementary to up to one tandem repeat of the DNA tandem nucleotide  
repeat locus.
- 15 74. The kit of claim 71, wherein at least one of the plurality of primers comprises  
a sequence complementary to up to two tandem repeats of the DNA tandem  
nucleotide repeat locus.
- 20 75. The kit of claim 71, wherein at least one of the plurality of primers comprises  
a sequence complementary to up to three tandem repeats of the DNA tandem  
nucleotide repeat locus.

76. The kit of claim 71, wherein the extension of at least one primer is terminated using a chain termination reagent.
77. The kit of claim 71, wherein the chain termination reagent is a dideoxynucleotide triphosphate.
78. The kit of claim 71, wherein at least one of the plurality of primers comprises a cleavable site.
79. The kit of claim 78, wherein the cleavable site comprises a recognition site for a restriction endonuclease, an exonuclease blocking site, or a chemically cleavable site.
80. The kit of claim 79, wherein the cleavable site comprises a recognition site for a restriction endonuclease.
81. The kit of claim 79, wherein the cleavable site comprises a chemically cleavable site.
82. The kit of claim 78, wherein at least one of the plurality of primers is capable of attaching to a solid support.
83. The kit of claim 82, wherein at least one of the plurality of primers comprises biotin or digoxigenin.
84. The kit of claim 71 wherein at least one of the DNA tandem nucleotide repeat loci is selected from the group consisting of CSF1PO, D5S818, D7S820, D8S1179, D13S317, D16S539, D16S539, D18S51, D21S11, DYS19, F13A1, FES/FPS, FGA, HPRTB, TH01, TPOX, DYS388, DYS391, DYS392, DYS393, D2S1391, D18S535, D2S1338, D19S433, D6S477, D1S518, D14S306, D22S684, F13B, CD4, D12S391, D10S220 and D7S523.

85. The kit of claim 84, wherein two of the plurality of primers comprise a pair of sequences selected from the group consisting of

a sequence according to SEQ ID NO:1 and a sequence according to SEQ ID NO:2;

a sequence according to SEQ ID NO:3 and a sequence according to SEQ ID NO:4;

a sequence according to SEQ ID NO:5 and a sequence according to SEQ ID NO:6;

a sequence according to SEQ ID NO:7 and a sequence according to SEQ ID NO:8;

a sequence according to SEQ ID NO:9 and a sequence according to SEQ ID NO:10;

a sequence according to SEQ ID NO:11 and a sequence according to SEQ ID NO:12;

a sequence according to SEQ ID NO:13 and a sequence according to SEQ ID NO:14;

a sequence according to SEQ ID NO:15 and a sequence according to SEQ ID NO:16;

a sequence according to SEQ ID NO:17 and a sequence according to SEQ ID NO:18;

a sequence according to SEQ ID NO:19 and a sequence according to SEQ ID NO:20;

a sequence according to SEQ ID NO:21 and a sequence according to SEQ ID NO:22;

a sequence according to SEQ ID NO:23 and a sequence according to SEQ ID NO:24;

a sequence according to SEQ ID NO:25 and a sequence according to SEQ ID NO:26;

a sequence according to SEQ ID NO:27 and a sequence according to SEQ ID NO:28;

a sequence according to SEQ ID NO:29 and a sequence according to SEQ ID NO:30;

a sequence according to SEQ ID NO:31 and a sequence according to SEQ ID NO:32;

a sequence according to SEQ ID NO:49 and a sequence according to SEQ ID NO:83;

a sequence according to SEQ ID NO:52 and a sequence according to SEQ ID NO:84;

a sequence according to SEQ ID NO:54 and a sequence according to SEQ ID NO:85;

a sequence according to SEQ ID NO:56 and a sequence according to SEQ ID NO:86;

a sequence according to SEQ ID NO:58 and a sequence according to SEQ ID NO:87;

a sequence according to SEQ ID NO:59 and a sequence according to SEQ ID NO:88;

a sequence according to SEQ ID NO:62 and a sequence according to SEQ ID NO:89;

a sequence according to SEQ ID NO:63 and a sequence according to SEQ ID NO:90;

a sequence according to SEQ ID NO:66 and a sequence according to SEQ ID NO:91;

a sequence according to SEQ ID NO:67 and a sequence according to SEQ ID NO:92;

a sequence according to SEQ ID NO:70 and a sequence according to SEQ ID NO:93;

a sequence according to SEQ ID NO:72 and a sequence according to SEQ ID NO:94;

a sequence according to SEQ ID NO:74 and a sequence according to SEQ ID NO:95;

a sequence according to SEQ ID NO:76 and a sequence according to SEQ ID NO:96;

a sequence according to SEQ ID NO:78 and a sequence according to SEQ ID NO:97;

a sequence according to SEQ ID NO:80 and a sequence according to SEQ ID NO:98;

a sequence according to SEQ ID NO:66 and a sequence according to SEQ ID NO:99;

a sequence according to SEQ ID NO:33 and a sequence according to SEQ ID NO:100 and

a sequence according to SEQ ID NO:101 and a sequence according to SEQ ID NO:103.

86. An oligonucleotide primer comprising a nucleic acid sequence complementary to a sequence flanking a DNA tandem nucleotide repeat region, wherein extension of the primer generates nucleic acid products for mass spectral analysis, said nucleic acid products having 10 to 100 nucleotides in length.

87. The method of claim 86, wherein a 3' end of one or more primers immediately flanks a DNA tandem nucleotide repeat region.

88. The oligonucleotide primer of claim 86, further comprising a sequence complementary to up to one tandem repeat of the DNA tandem nucleotide repeat locus.

89. The oligonucleotide primer of claim 86, further comprising a sequence complementary to up to two tandem repeats of the DNA tandem nucleotide repeat locus.

90. The oligonucleotide primer of claim 86, further comprising a sequence complementary to up to three tandem repeats of the DNA tandem nucleotide repeat locus.
- 5 91. The oligonucleotide primer of claim 86 wherein the locus is selected from the group consisting of CSF1PO, D5S818, D7S820, D8S1179, D13S317, D16S539, D16S539, D18S51, D21S11, DYS19, F13A1, FES/FPS, FGA, HPRTB, TH01, TPOX, DYS388, DYS391, DYS392, DYS393, D2S1391, D18S535, D2S1338, D19S433, D6S477, D1S518, D14S306, D22S684, F13B, CD4, D12S391, D10S220  
10 and D7S523.
92. The primer of claim 86, wherein the sequence flanking the DNA tandem nucleotide repeat locus comprises a flanking sequence selected from the group consisting of a sequence according to SEQ ID NO:104, a sequence according to SEQ ID NO:105, a sequence according to SEQ ID NO:106, a sequence according to SEQ ID NO:107, a sequence according to SEQ ID NO:108, a sequence according to SEQ ID NO:109, a sequence according to SEQ ID NO:110, a sequence according to SEQ ID NO:111, a sequence according to SEQ ID NO:112, a sequence according to SEQ ID NO:113, a sequence according to SEQ ID NO:114, a sequence according to SEQ ID NO:115, a sequence according to SEQ ID NO:116, a sequence according to SEQ ID NO:117, a sequence according to SEQ ID NO:118, a sequence according to SEQ ID NO:119, a sequence according to SEQ ID NO:120, a sequence according to SEQ ID NO:121, a sequence according to SEQ ID NO:122, a sequence according to SEQ ID NO:123, a sequence according to SEQ ID NO:124, a sequence according to SEQ ID NO:125, a sequence according to SEQ ID NO:126, a sequence according to SEQ ID NO:127, a sequence according to SEQ ID NO:128, a sequence according to SEQ ID NO:129, a sequence according to SEQ ID NO:130, a sequence according to SEQ ID NO:131, a sequence according to SEQ ID NO:132, a sequence according to SEQ ID NO:133, a sequence according to SEQ ID NO:134, and a sequence according to SEQ ID NO:135.

93. The method of claim 92, wherein a 3' end of one or more primers immediately flanks a DNA tandem nucleotide repeat region.
94. The oligonucleotide primer of claim 92, further comprising a sequence  
5 complementary to up to one tandem repeat of the DNA tandem nucleotide repeat locus.
95. The oligonucleotide primer of claim 92, further comprising a sequence complementary to up to two tandem repeats of the DNA tandem nucleotide repeat locus.
96. The oligonucleotide primer of claim 92, wherein the primer further comprises  
10 a sequence complementary to up to three tandem repeats of the DNA tandem nucleotide repeat locus.
97. The oligonucleotide primer of claim 86, wherein the oligonucleotide primer further comprises a cleavable site.
98. The oligonucleotide primer of claim 97, wherein the cleavable site comprises a recognition site for a restriction endonuclease, an exonuclease blocking site, or a chemically cleavable site.
99. The oligonucleotide primer of claim 91, wherein the oligonucleotide primer comprises a nucleotide sequence selected from the group consisting of a sequence according to SEQ ID NO:1, a sequence according to SEQ ID NO:2, a sequence according to SEQ ID NO:3, a sequence according to SEQ ID NO:4, a sequence according to SEQ ID NO:5, a sequence according to SEQ ID NO:6, a sequence according to SEQ ID NO:7, a sequence according to SEQ ID NO:8, a sequence according to SEQ ID NO:9, a sequence according to SEQ ID NO:10, a sequence according to SEQ ID NO:11, a sequence according to SEQ ID NO:12, a sequence according to SEQ ID NO:13, a sequence according to SEQ ID NO:14, a sequence

[illegible]



according to SEQ ID NO:77, a sequence according to SEQ ID NO:78, a sequence according to SEQ ID NO:79, a sequence according to SEQ ID NO:80, a sequence according to SEQ ID NO:81, a sequence according to SEQ ID NO:82, a sequence according to SEQ ID NO:83, a sequence according to SEQ ID NO:84, a sequence according to SEQ ID NO:85, a sequence according to SEQ ID NO:86, a sequence according to SEQ ID NO:87, a sequence according to SEQ ID NO:88, a sequence according to SEQ ID NO:89, a sequence according to SEQ ID NO:90, a sequence according to SEQ ID NO:91, a sequence according to SEQ ID NO:92, a sequence according to SEQ ID NO:93, a sequence according to SEQ ID NO:94, a sequence according to SEQ ID NO:95, a sequence according to SEQ ID NO:96, a sequence according to SEQ ID NO:97, a sequence according to SEQ ID NO:98, a sequence according to SEQ ID NO:99, a sequence according to SEQ ID NO:100, a sequence according to SEQ ID NO:101, a sequence according to SEQ ID NO:102 and a sequence according to SEQ ID NO:103.

100. The oligonucleotide sequence of claim 99, wherein the oligonucleotide primer further comprises a cleavable site.

101. The oligonucleotide primer of claim 100, wherein the oligonucleotide primer further comprises biotin or digoxigenin.

102. The oligonucleotide sequence of claim 99, wherein the oligonucleotide sequence is selected from the group consisting of a sequence according to SEQ ID NO:2, a sequence according to SEQ ID NO:4, a sequence according to SEQ ID NO:5, a sequence according to SEQ ID NO:7, a sequence according to SEQ ID NO:9, a  
 5 sequence according to SEQ ID NO:11, a sequence according to SEQ ID NO:14, a sequence according to SEQ ID NO:16, a sequence according to SEQ ID NO:17, a sequence according to SEQ ID NO:19, a sequence according to SEQ ID NO:21, a sequence according to SEQ ID NO:23, a sequence according to SEQ ID NO:25, a sequence according to SEQ ID NO:27, a sequence according to SEQ ID NO:30, a  
 10 sequence according to SEQ ID NO:31, a sequence according to SEQ ID NO:83, a

sequence according to SEQ ID NO:84, a sequence according to SEQ ID NO:85, a  
sequence according to SEQ ID NO:86, a sequence according to SEQ ID NO:87, a  
sequence according to SEQ ID NO:88, a sequence according to SEQ ID NO:89, a  
sequence according to SEQ ID NO:90, a sequence according to SEQ ID NO:91, a  
5 sequence according to SEQ ID NO:92, a sequence according to SEQ ID NO:93, a  
sequence according to SEQ ID NO:94, a sequence according to SEQ ID NO:95, a  
sequence according to SEQ ID NO:96, a sequence according to SEQ ID NO:97, a  
sequence according to SEQ ID NO:98, a sequence according to SEQ ID NO:99, a  
sequence according to SEQ ID NO:100, a sequence according to SEQ ID NO:101, a  
10 sequence according to SEQ ID NO:102, and a sequence according to SEQ ID  
NO:103.

103. An oligonucleotide primer comprising a nucleic acid sequence complementary  
to a sequence flanking a DNA tandem nucleotide repeat locus and a cleavable site,  
15 wherein extension of the primer generates a limited size range of nucleic acid products  
for mass spectral analysis.

104. The oligonucleotide primer of claim 103, wherein the cleavable site comprises  
a recognition site for a restriction endonuclease, an exonuclease blocking site, or a  
chemically cleavable site.

105. The oligonucleotide primer of claim 103, further comprising a sequence  
complementary to up to one tandem repeat of the DNA tandem nucleotide repeat  
locus.

20

106. The oligonucleotide primer of claim 103, further comprising a sequence  
complementary to up to two tandem repeats of the DNA tandem nucleotide repeat  
locus.

107. The oligonucleotide primer of claim 103, further comprising a sequence complementary to up to three tandem repeats of the DNA tandem nucleotide repeat locus.
- 5 108. The oligonucleotide primer of claim 103 wherein the locus is selected from the group consisting of CSF1PO, D5S818, D7S820, D8S1179, D13S317, D16S539, D16S539, D18S51, D21S11, DYS19, T13A1, FES/FPS, FGA, JPRTR, TH01, TPOX, DYS388, DYS391, DYS392, DYS393, D2S1391, D18S535, D2S1338, D19S433, D6S477, D1S518, D14S306, D22S684, F13B, CD4, D12S391, D10S220  
10 and D7S523.
109. The primer of claim 103, wherein the sequence flanking the DNA tandem nucleotide repeat locus comprises a flanking sequence selected from the group consisting of a sequence according to SEQ ID NO:104, a sequence according to SEQ ID NO:105, a sequence according to SEQ ID NO:106, a sequence according to SEQ ID NO:107, a sequence according to SEQ ID NO:108, a sequence according to SEQ ID NO:109, a sequence according to SEQ ID NO:110, a sequence according to SEQ ID NO:111, a sequence according to SEQ ID NO:112, a sequence according to SEQ ID NO:113, a sequence according to SEQ ID NO:114, a sequence according to SEQ ID NO:115, a sequence according to SEQ ID NO:116, a sequence according to SEQ ID NO:117, a sequence according to SEQ ID NO:118, a sequence according to SEQ ID NO:119, a sequence according to SEQ ID NO:120, a sequence according to SEQ ID NO:121, a sequence according to SEQ ID NO:122, a sequence according to SEQ ID NO:123, a sequence according to SEQ ID NO:124, a sequence according to SEQ ID NO:125, a sequence according to SEQ ID NO:126, a sequence according to SEQ ID NO:127, a sequence according to SEQ ID NO:128, a sequence according to SEQ ID NO:129, a sequence according to SEQ ID NO:130, a sequence according to SEQ ID NO:131, a sequence according to SEQ ID NO:132, a sequence according to SEQ ID NO:133, a sequence according to SEQ ID NO:134, and a sequence according to SEQ ID NO:135.

110. The oligonucleotide primer of claim 109, further comprising a sequence complementary to up to one tandem repeat of the DNA tandem nucleotide repeat locus.

111. The oligonucleotide primer of claim 109, further comprising a sequence complementary to up to two tandem repeats of the DNA tandem nucleotide repeat locus.

5

112. The oligonucleotide primer of claim 109, wherein the primer further comprises a sequence complementary to up to three tandem repeats of the DNA tandem nucleotide repeat locus.

1/24

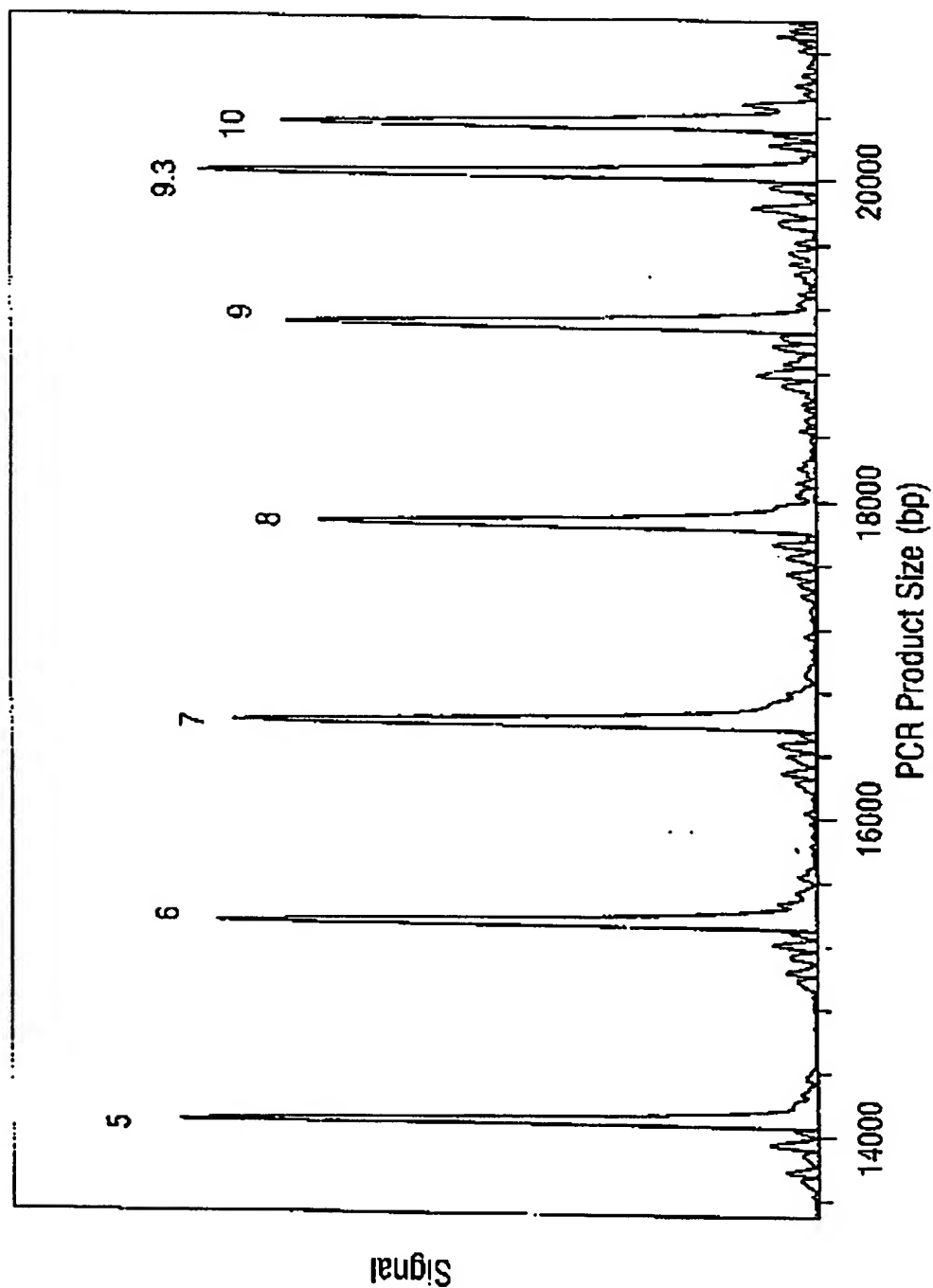
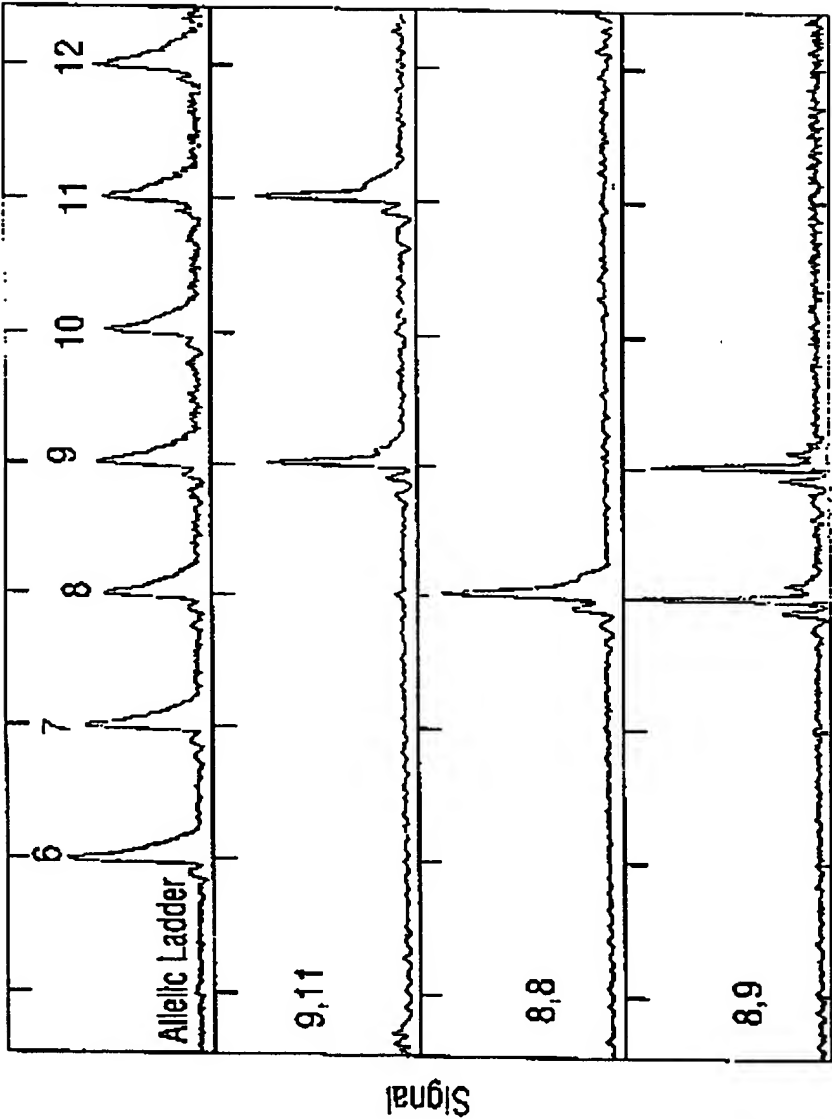


FIG. 1



Number of Repeats  
FIG. 2

3/24

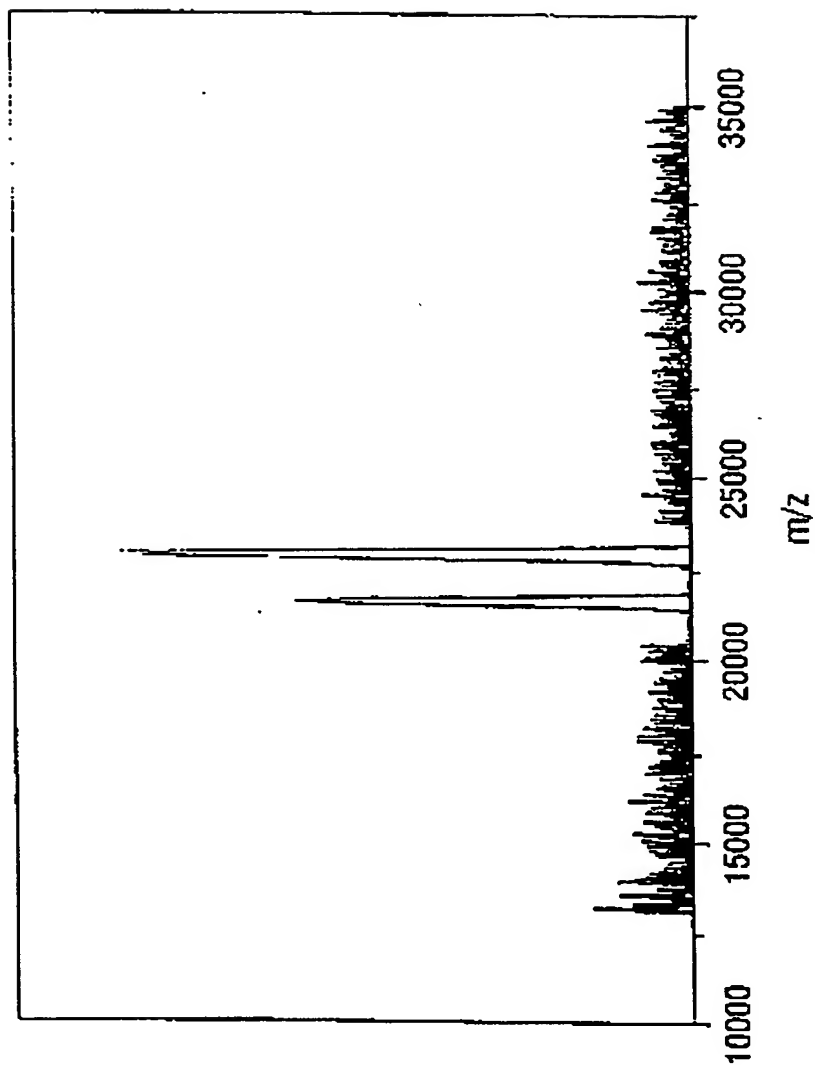


FIG. 3A

4/24

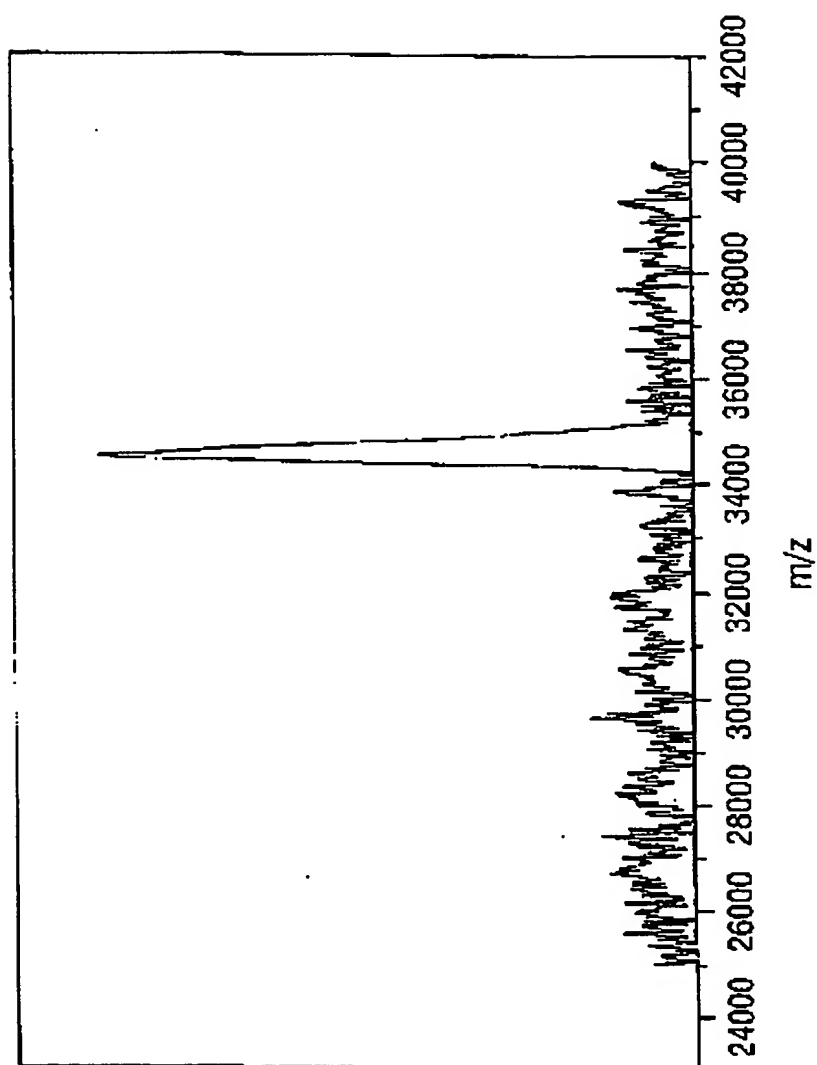


FIG. 3B



5/24

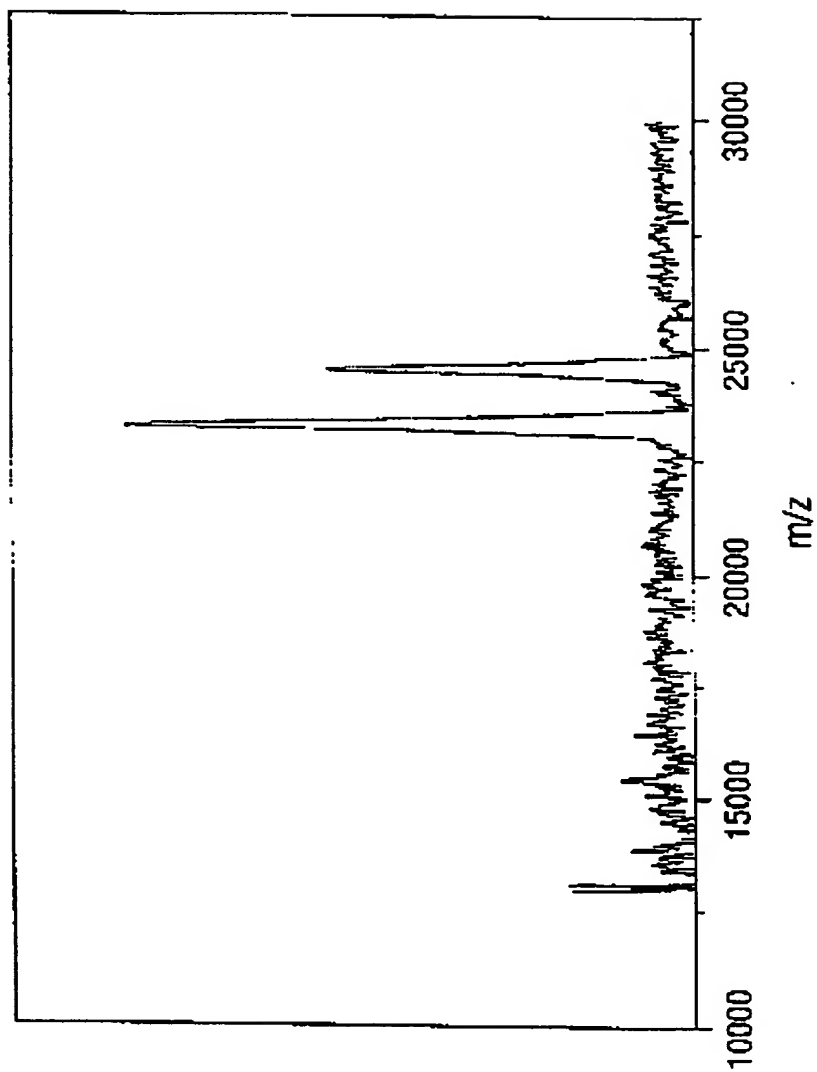


FIG. 3C

6/24

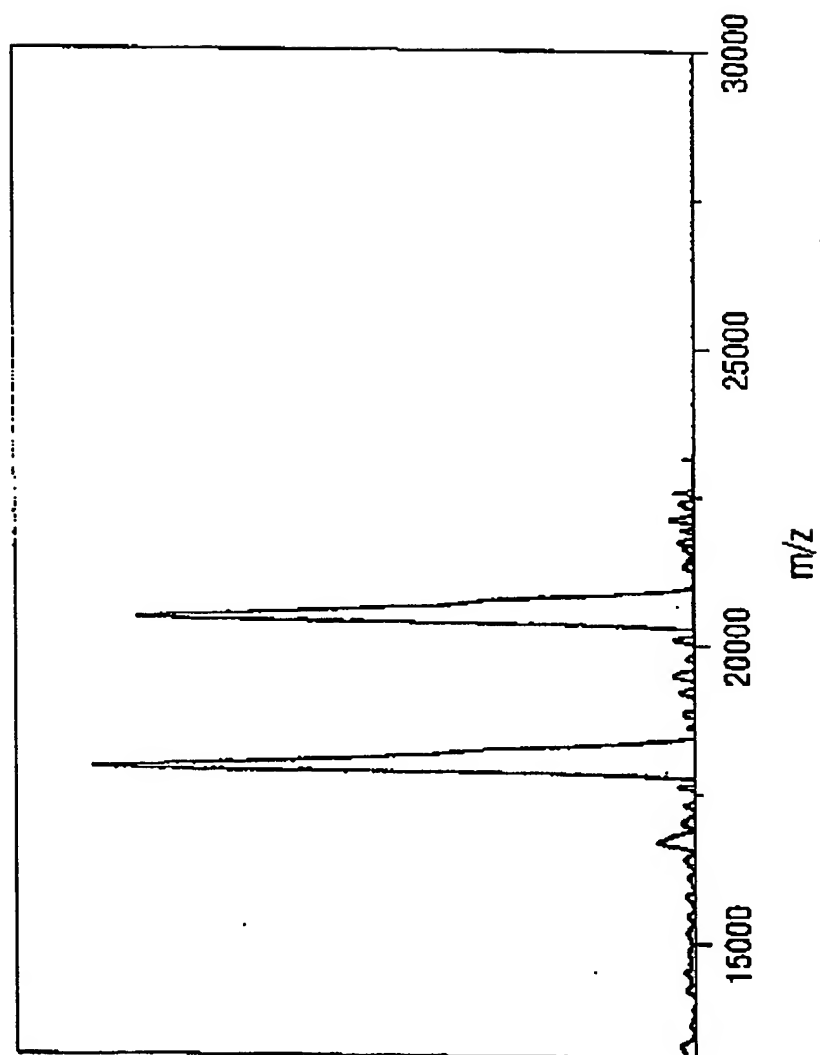


FIG. 3D

7/24

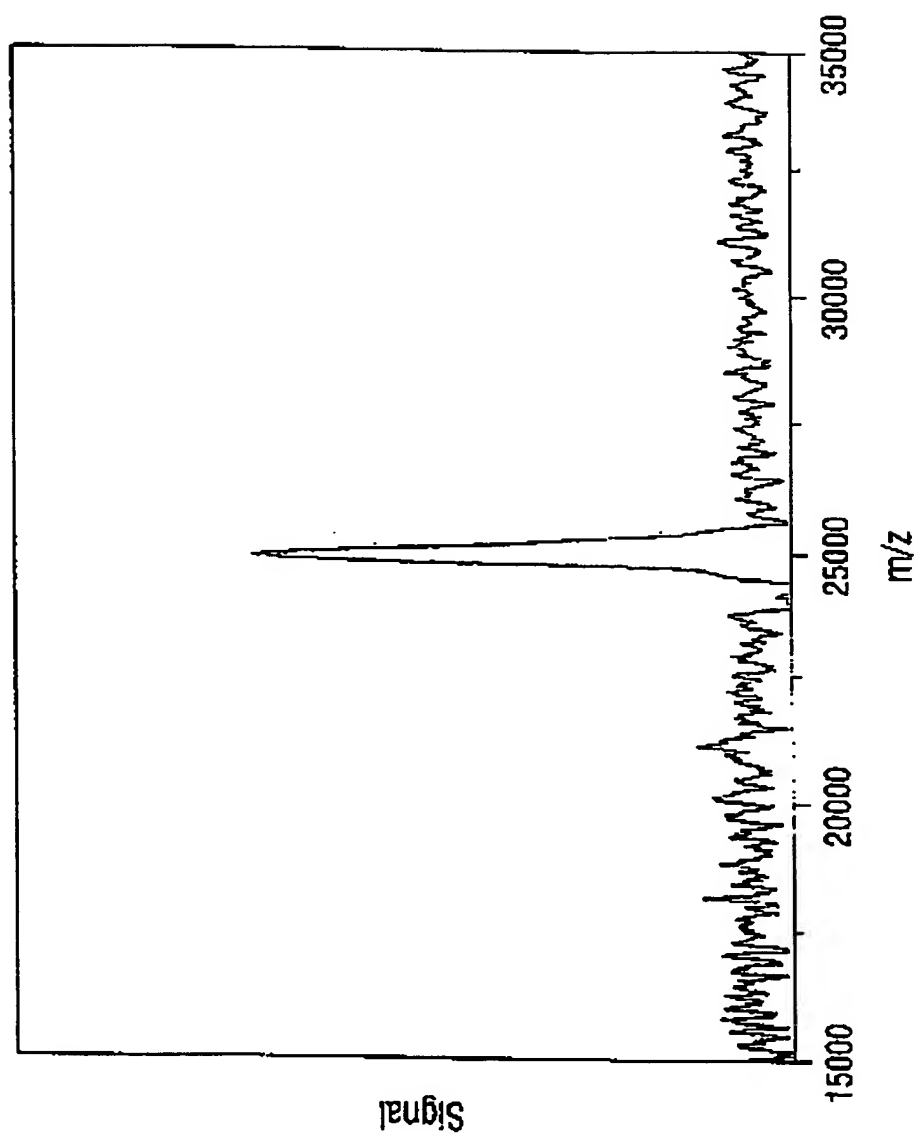


FIG. 3E

8/24

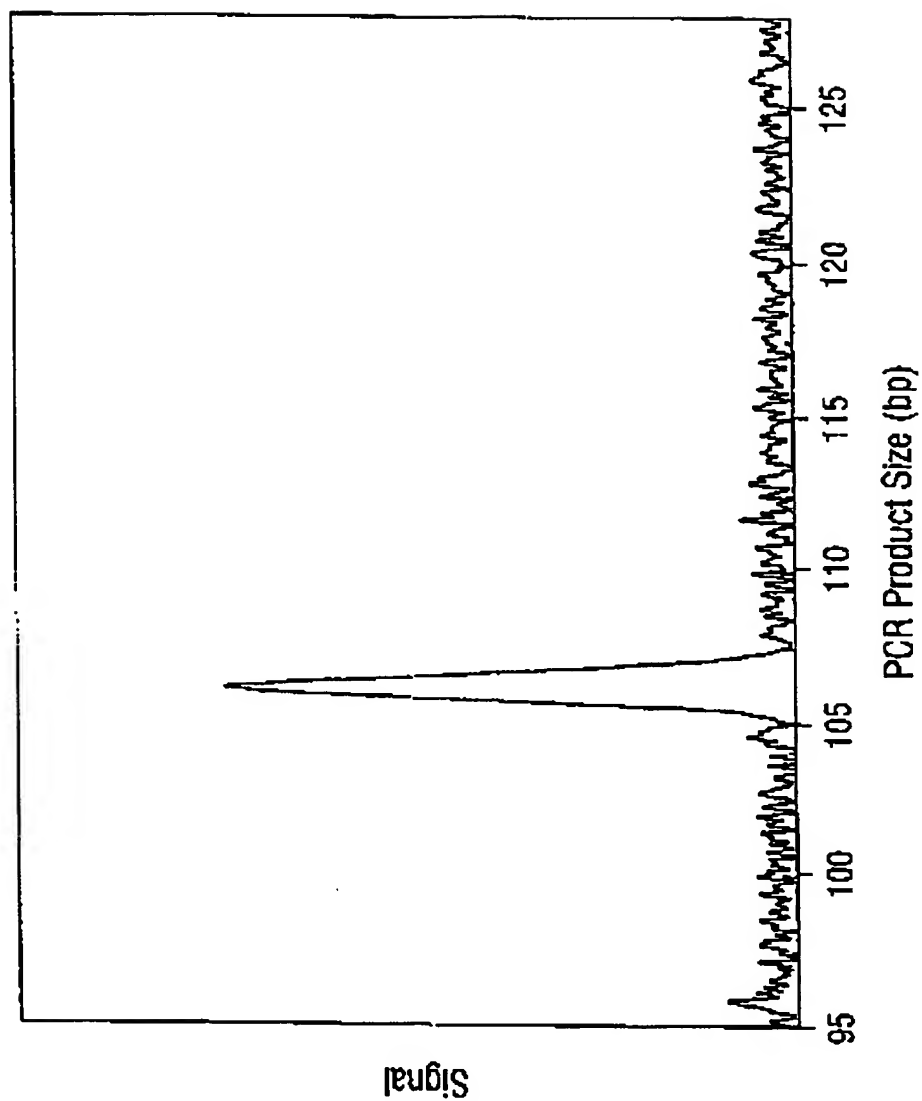


FIG. 3F

9/24

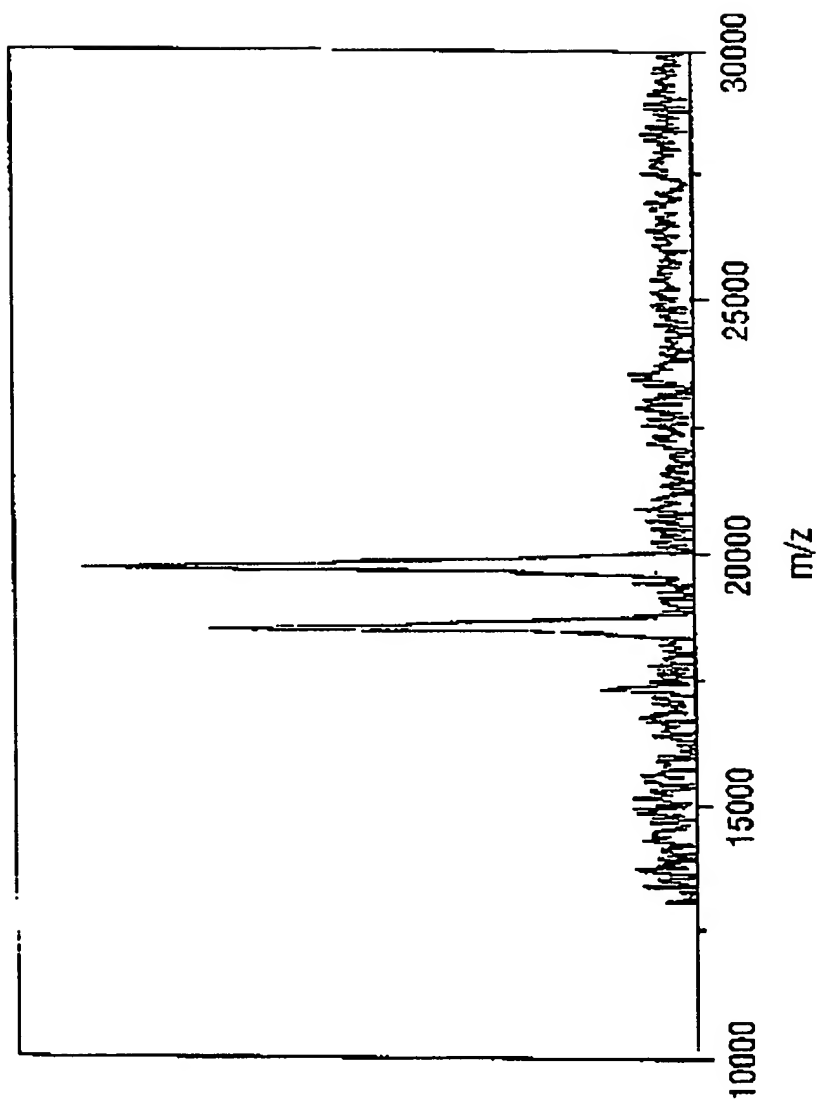


FIG. 3G

10/24

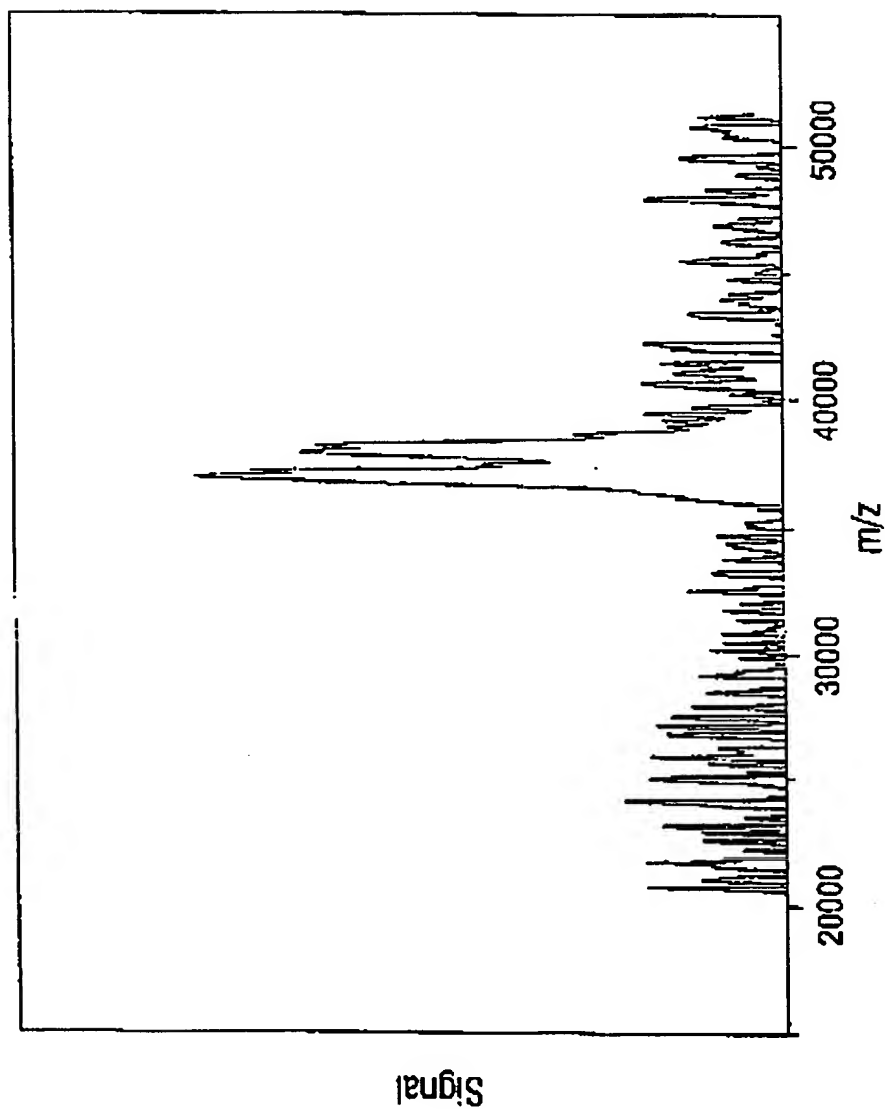


FIG. 3H

11/24

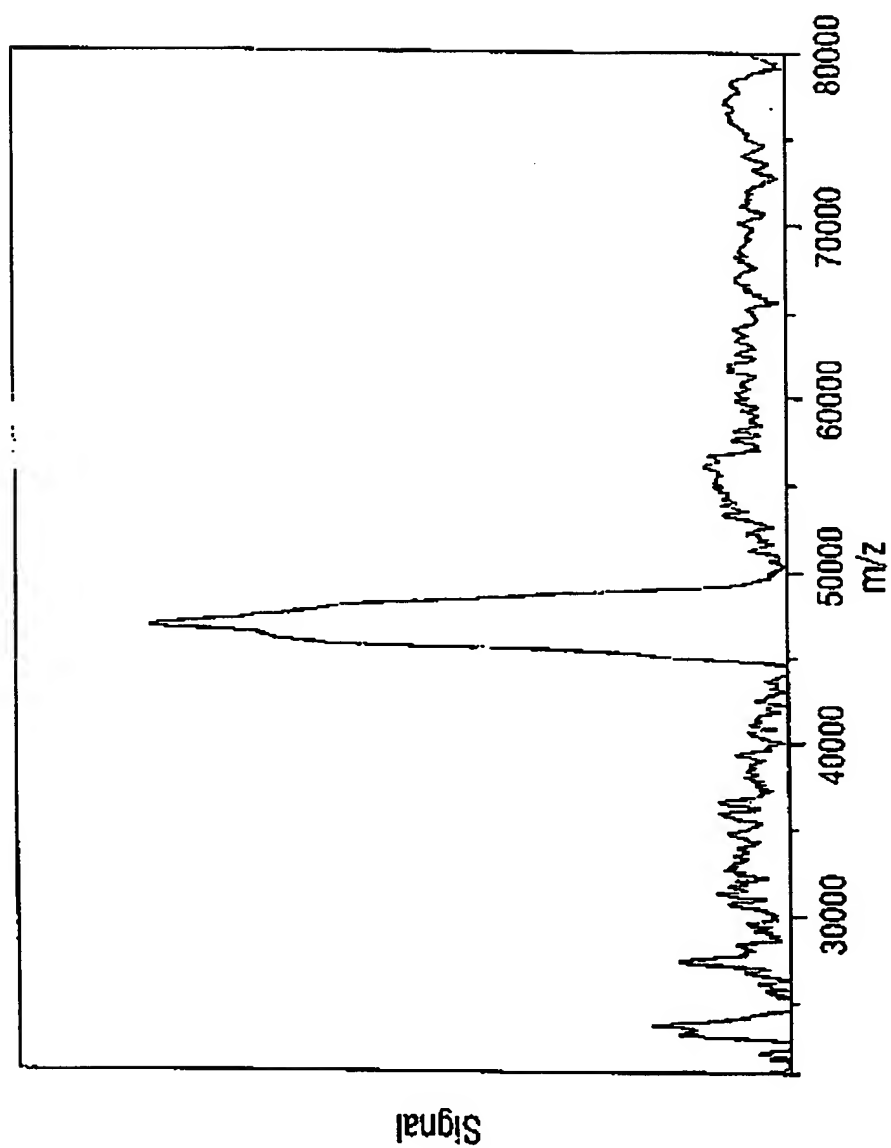


FIG. 31

12/24

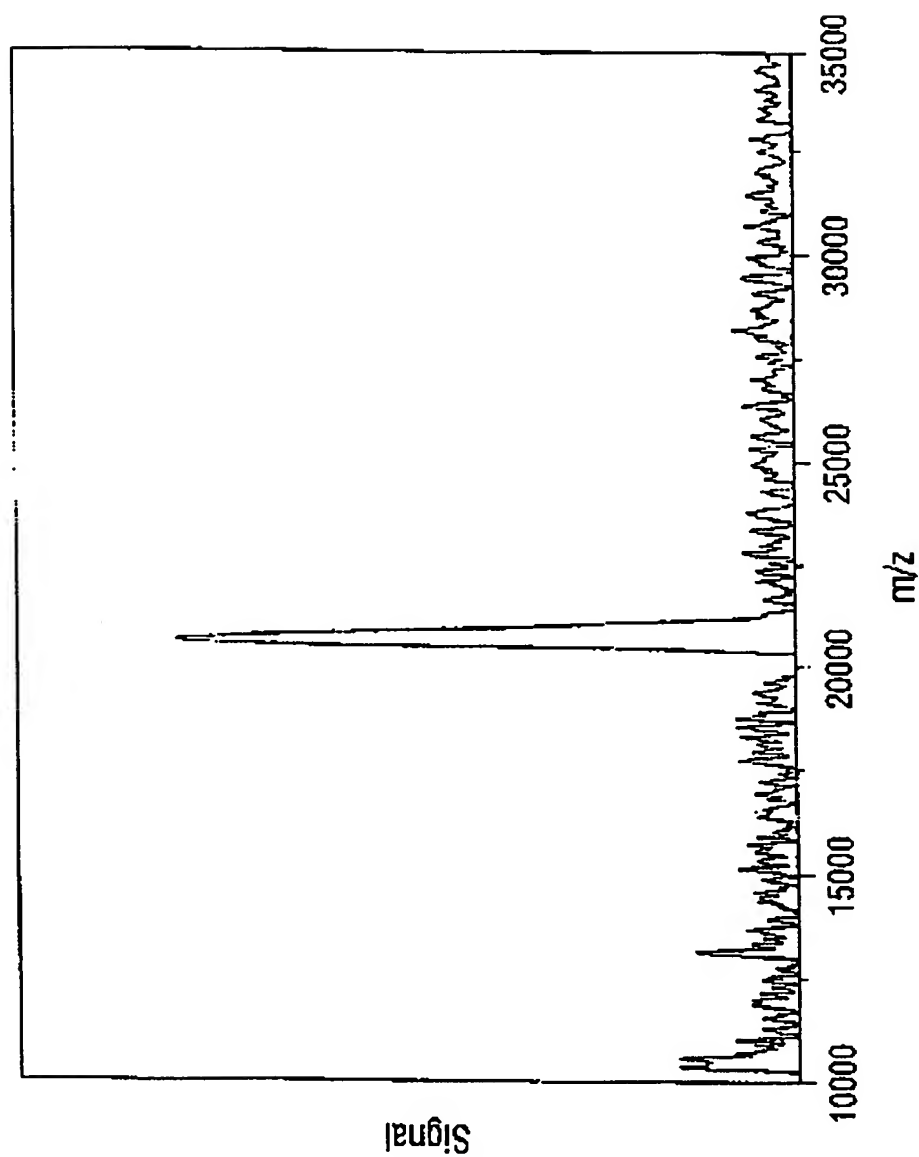


FIG. 3J



13/24

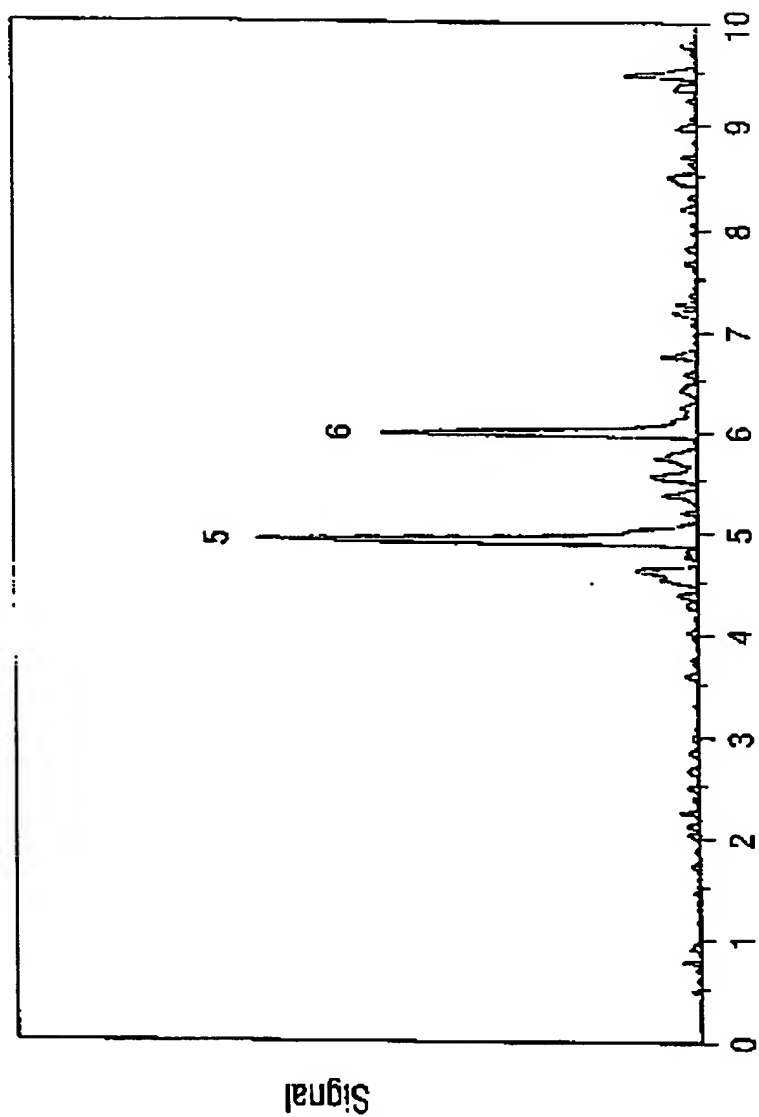
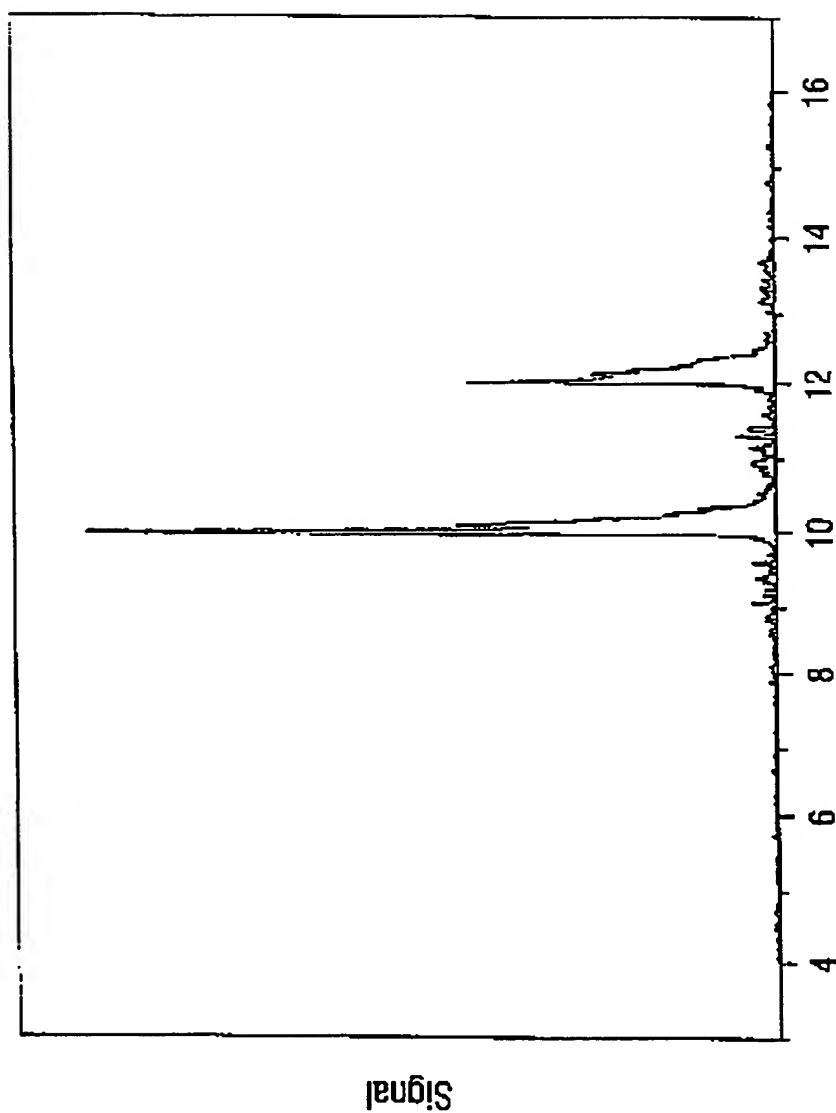


FIG. 3K

14/24

Repeats  
FIG. 3L

15/24

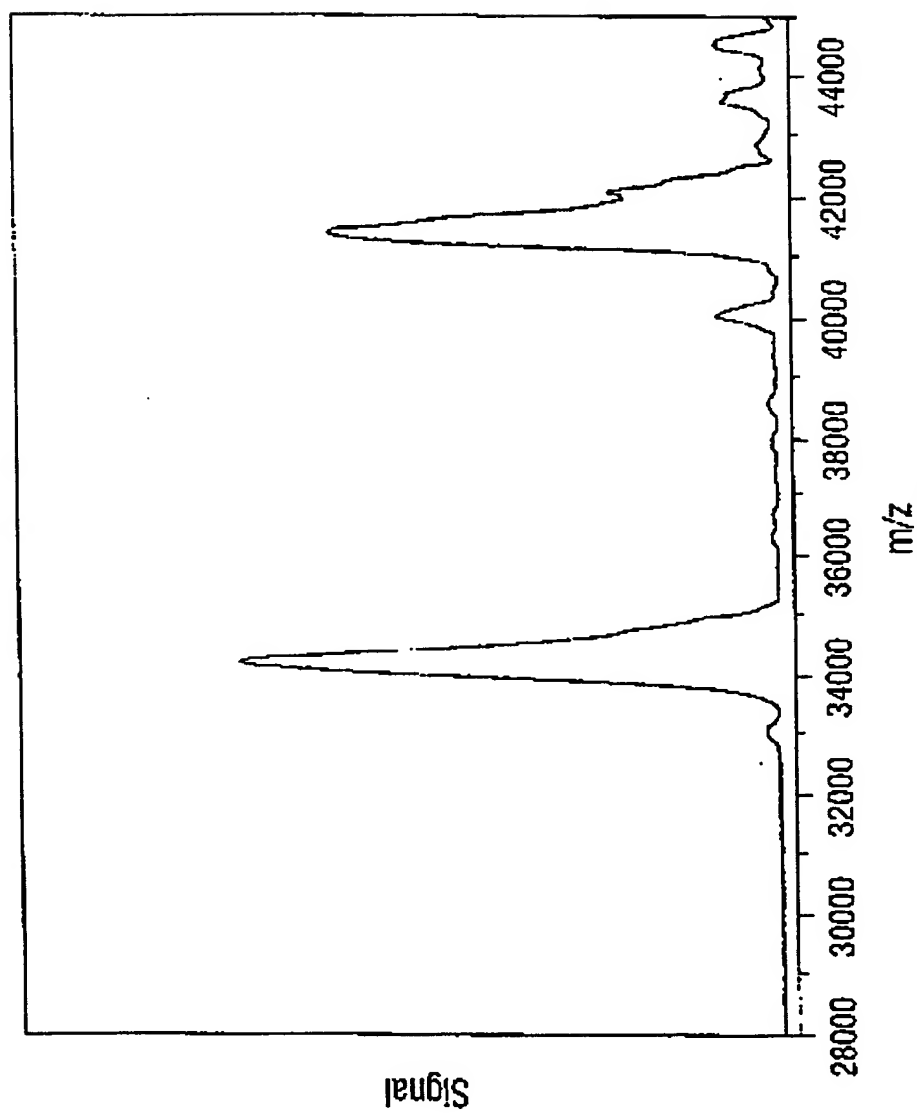


FIG. 3M

16/24

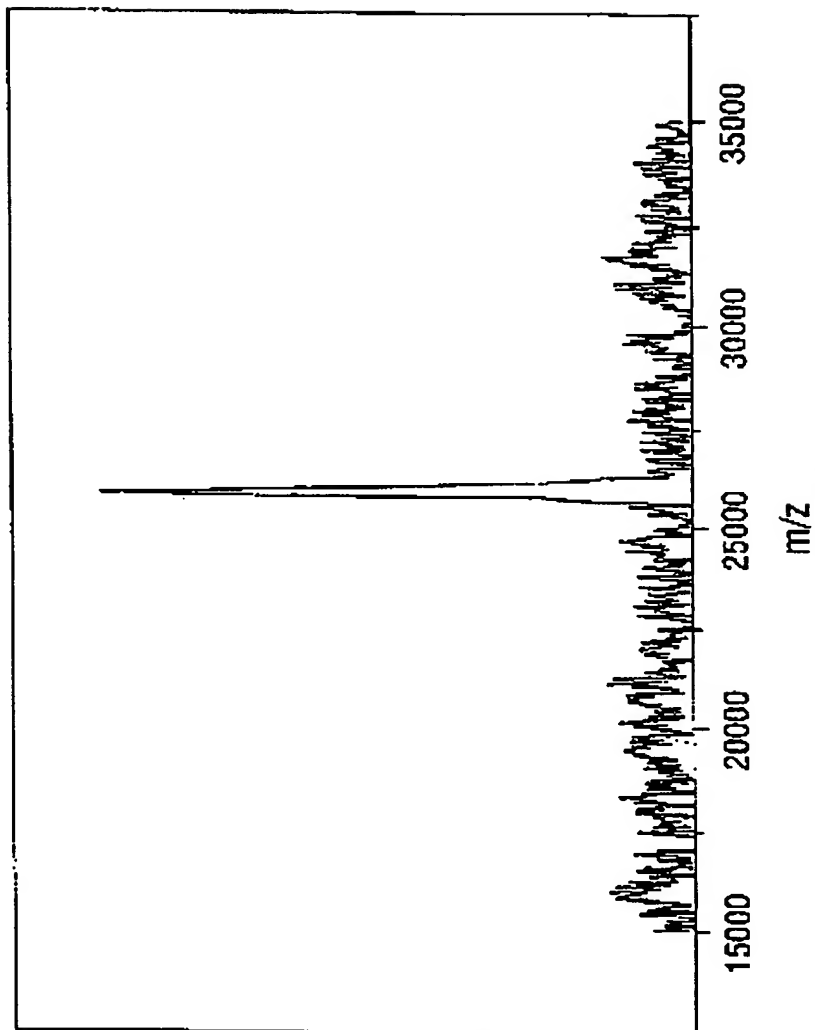


FIG. 3N

17/24

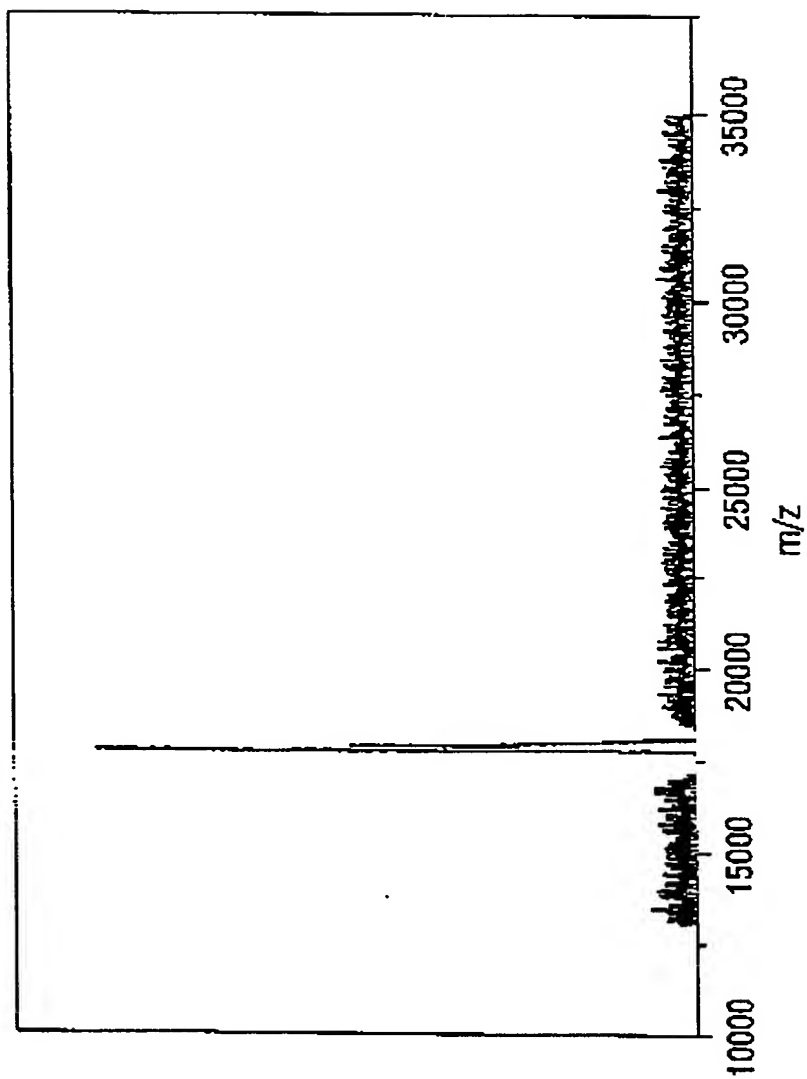


FIG. 30

18/24

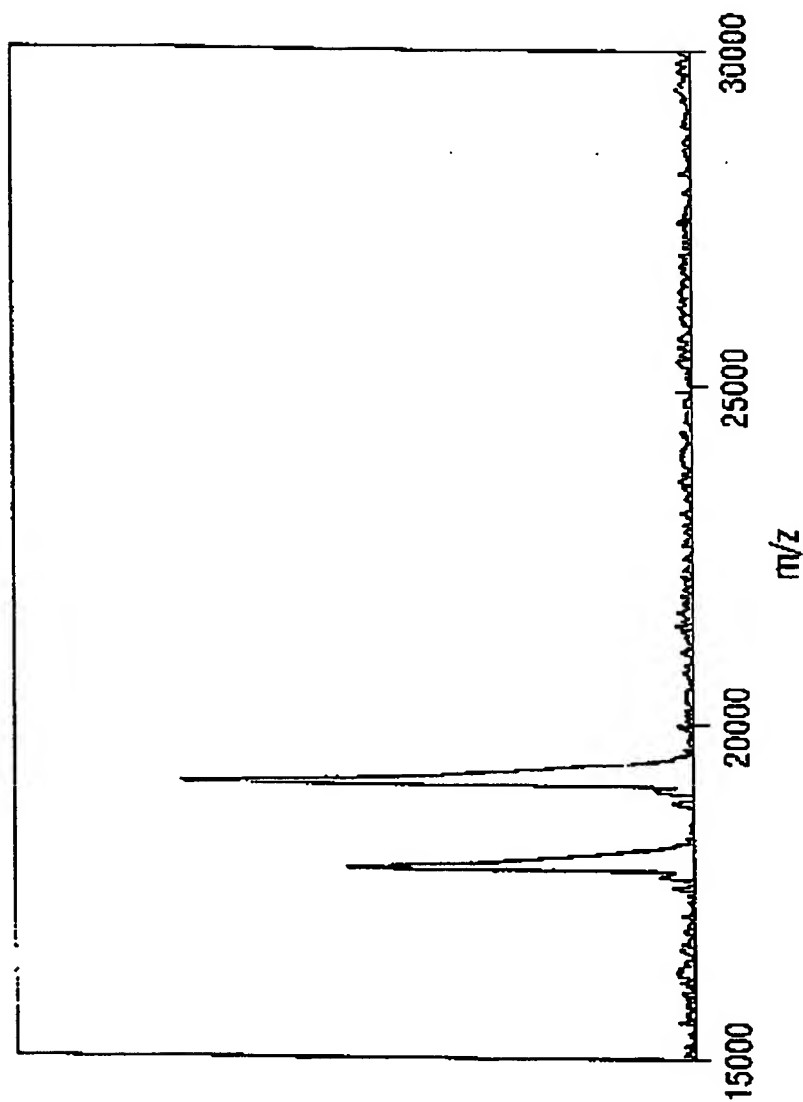


FIG. 3P

19/24

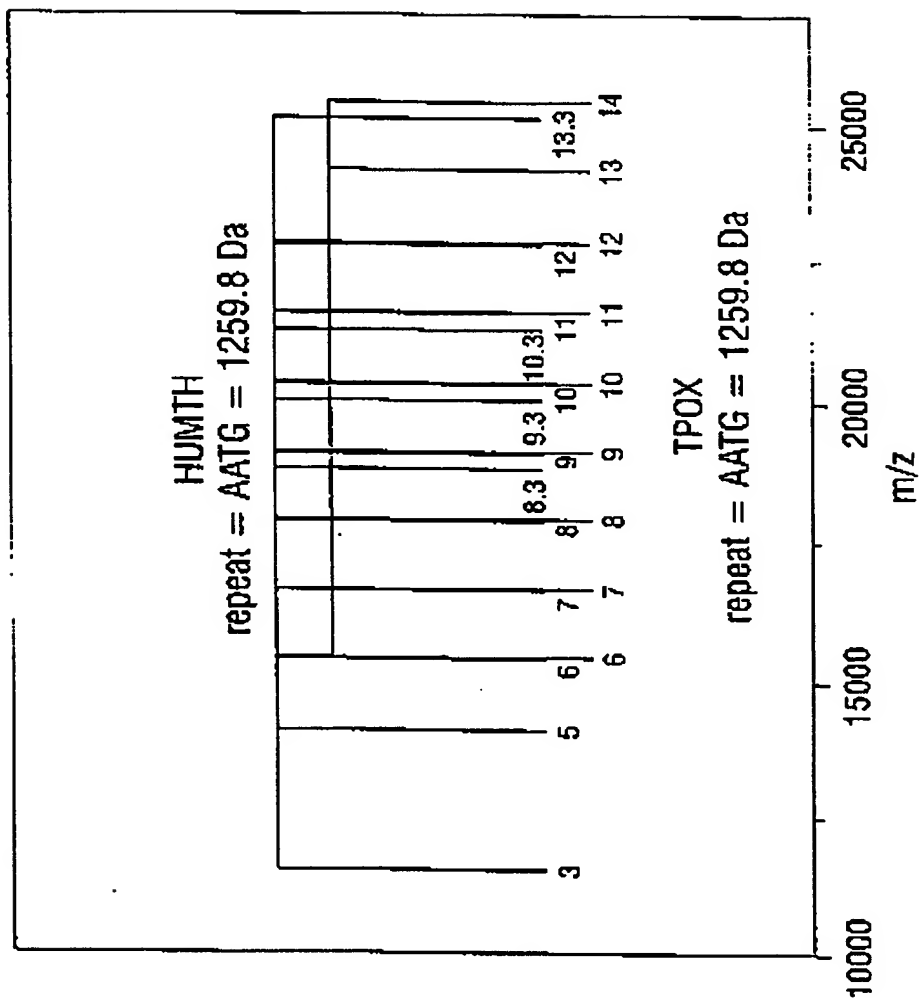
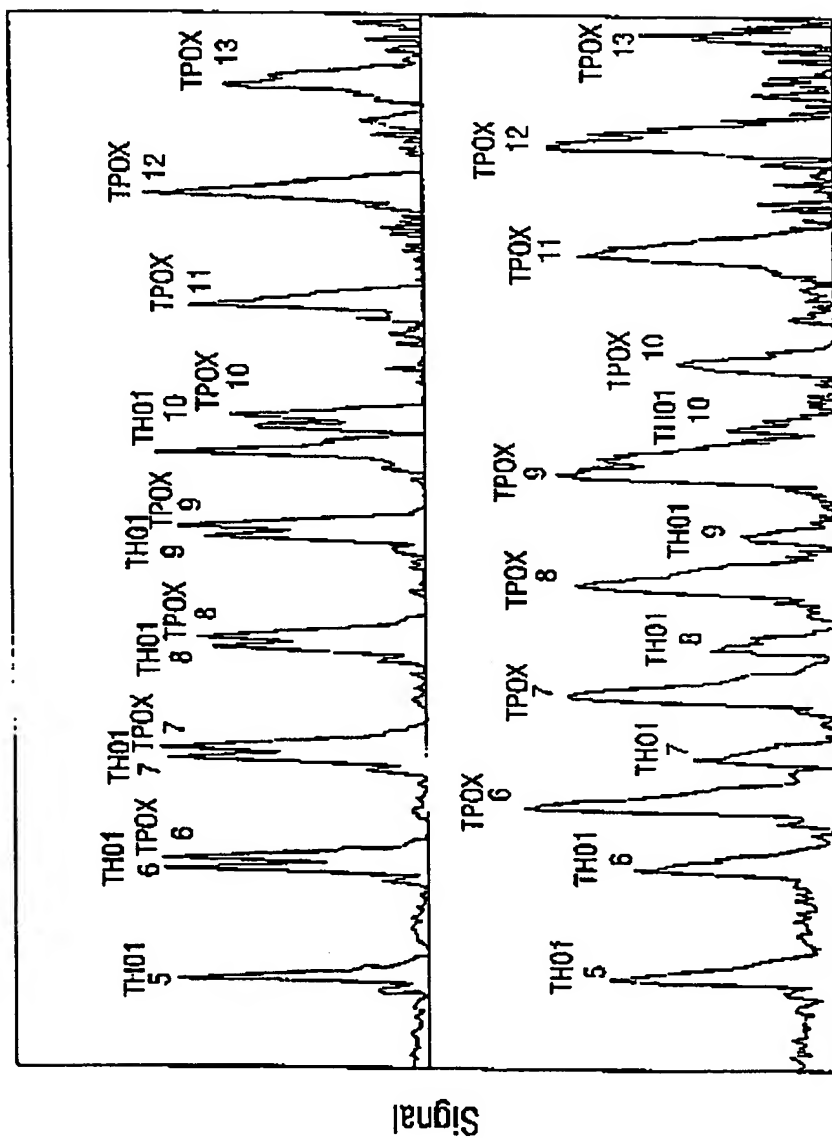


FIG. 4

20/24



Allele Designation

FIG. 5



21/24

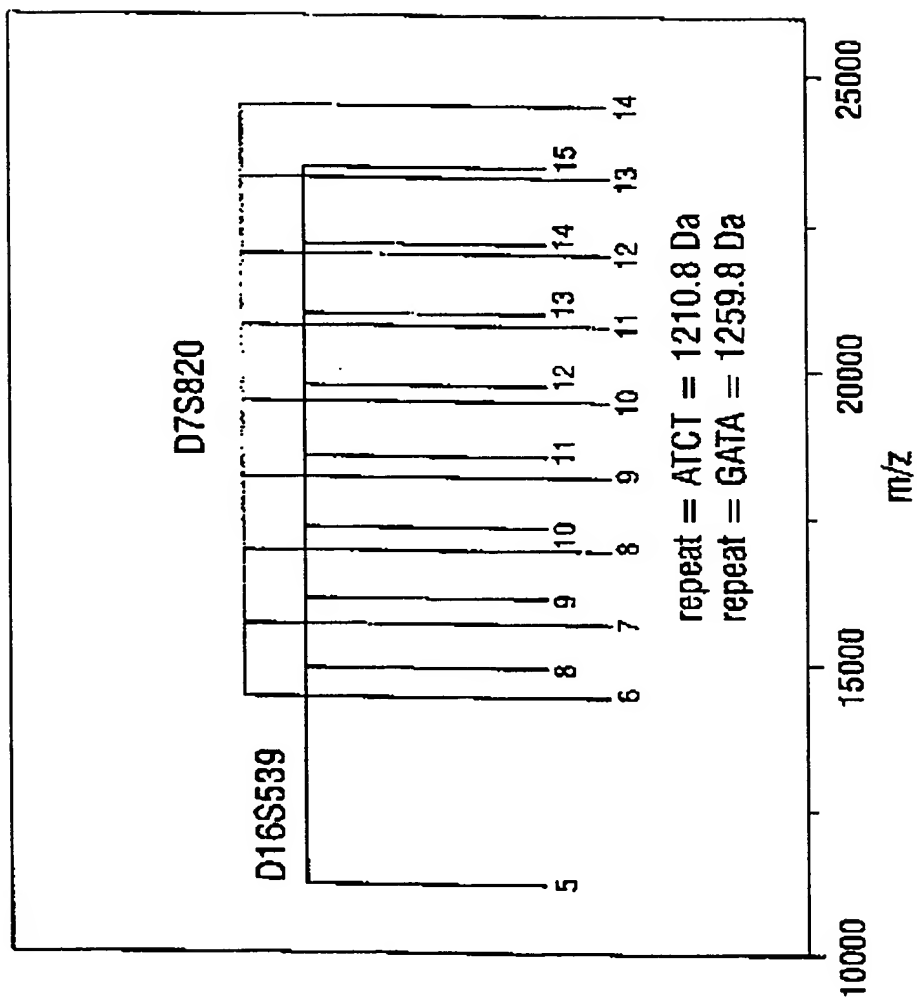


FIG. 6

22/24

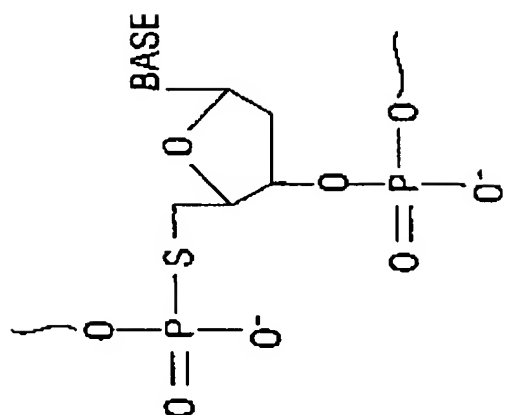


FIG. 7B

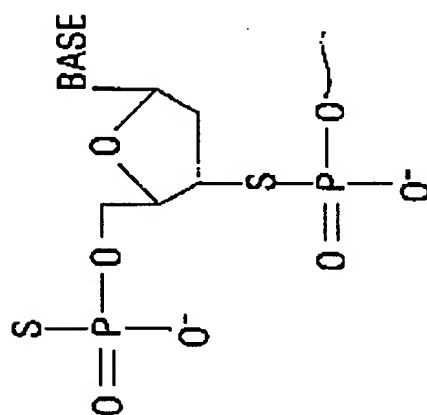
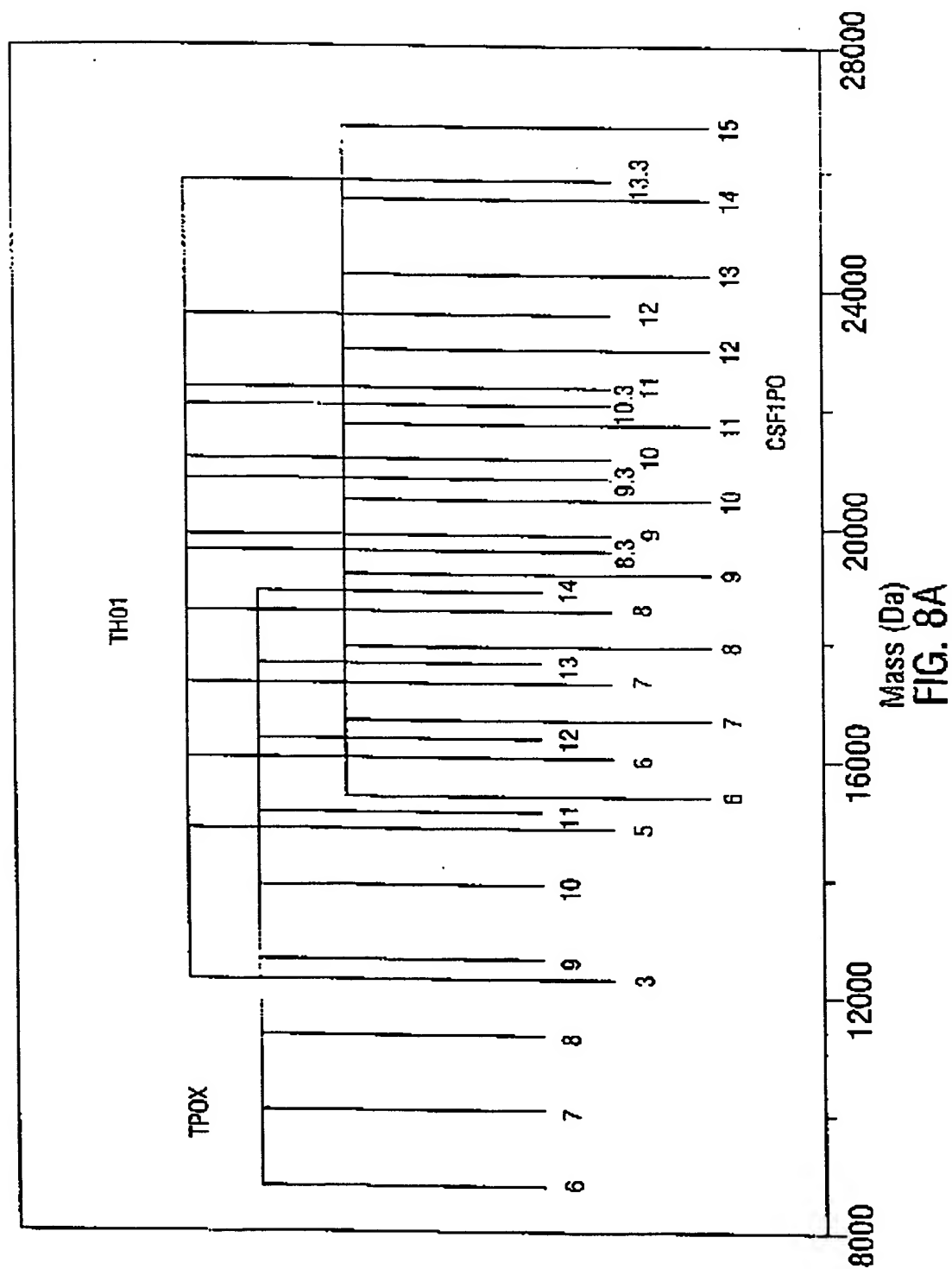


FIG. 7A

23/24



24/24

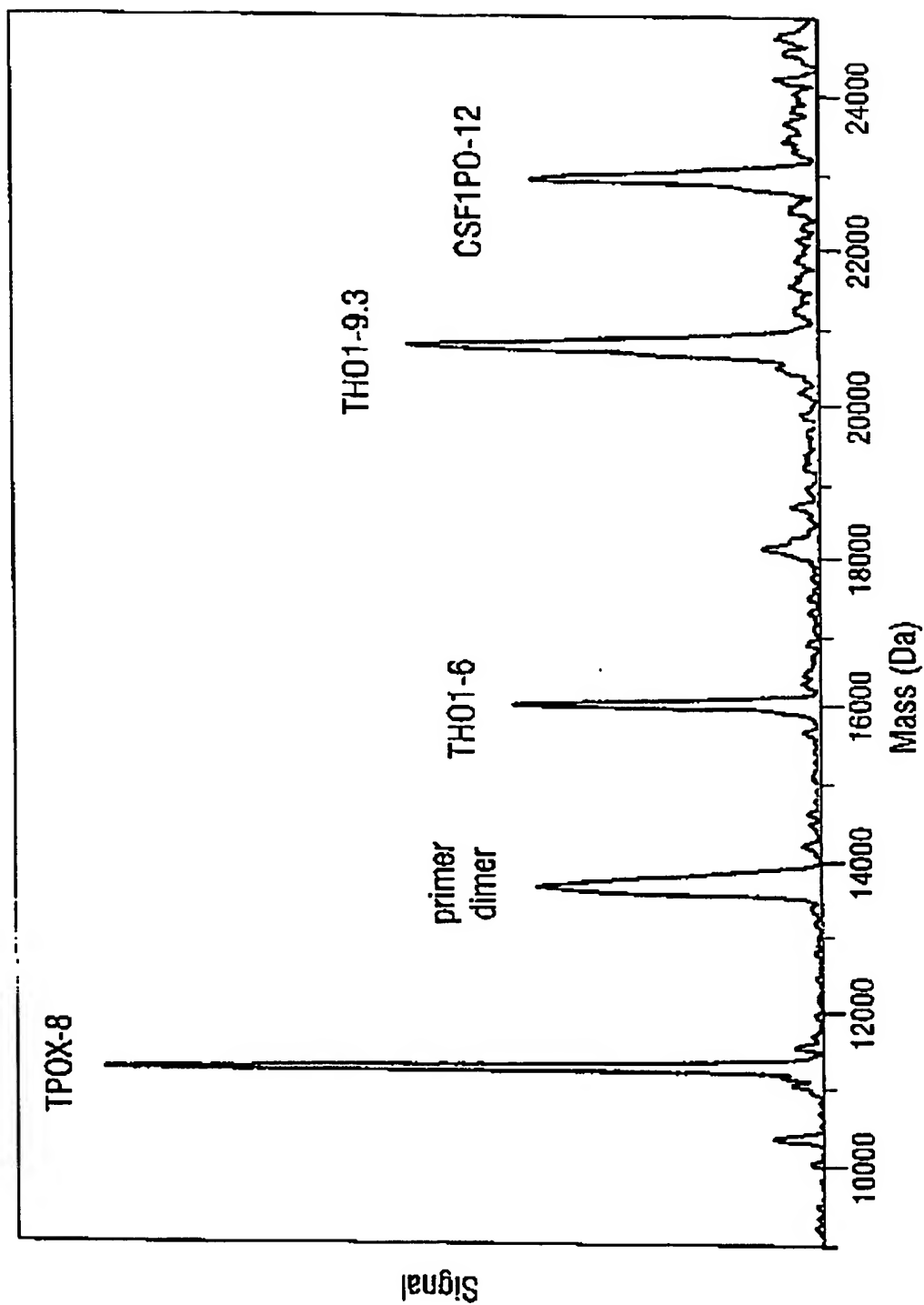


FIG. 8B

## SEQUENCE LISTING

<110> Butler, John M.

Li, Jia

Monforte, Joseph A.

Becker, Christopher H.

<210> DNA TYPING BY MASS SPECTROMETRY WITH POLYMORPHIC DNA  
REPEAT MARKERS

<130> GSTR:017/GSTR017P

<140> Unknown

<141> 1998-09-18

<150> 60/059,415

<151> 1997-09-19

<150> 135

<170> PatentIn Ver. 2.0

<210> 1

<211> 23

<212> DNA

<213> Homo sapiens

<400> 1

acagtaactg cctccataga tag

23

<210> 2

<211> 22

<212> DNA

<213> Homo sapiens

<220>

<231> misc\_feature

<222> (1)

<223> Biotinylated

<220>

<221> misc\_feature

<222> (21)

<223> 2'-deoxythymidine-5'-(3)-phosphorothioate

<400> 2

gtgtcagacc ctgtcttaag ta

22

<210> 3

<211> 19

<212> DNA

<213> Homo sapiens

<400> 2

actgcagtcac aatctgggt

29

<210> 4

<211> 20

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<222> (1)

<223> Biotinylated

<220>

<221> misc\_feature

<222> (19)

<223> 2'-deoxythymidine-5'-(5)-phosphorothioate

<400> 4

atgaaaluaa cagaggcttg

26

<210> 5

<211> 25

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<222> (1)

<223> Biotinylated

<220>

<221> misc\_feature

<222> (24)

<223> 2' deoxythymidino-5'-(3)-phosphorothioate

<400> 5

GCCTGGGTA CCTTATGTA ATATT

25

<210> 6

<211> 29

<212> DNA

<213> Homo sapiens

<400> 6

AAGGTACCC CATTATATC CTTATCTA

29

<210> 7

<211> 24

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<222> (1)

<223> Biotinylated

<220>

<221> misc\_feature

<222> (21)

<223> 2'-deoxythymidine-5'-phosphorothioate

<400> 7

tgccatagtt tagaacgaac taac

24

<210> 8

<211> 26

<212> DNA

<213> Homo sapiens

<400> 8

aaaactatc atctgtcta tctatc

26

<210> 9

<211> 27

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<222> (1)

<223> Biotinylated

<220>

<221> misc\_feature

<222> (26)

<223> 2' deoxythymidine-5'-(S)-phosphorothioate

<400> 9

ttgttatctc atg-gtacat tegtalc

27

<210> 10

<211> 28

<212> DNA

<213> Homo sapiens



&lt;400&gt; 10

acctatcctg tagattattt tcactgtg

28

&lt;210&gt; 11

&lt;211&gt; 23

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc feature

&lt;222&gt; (1)

&lt;223&gt; Sirtinylated

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (22)

&lt;223&gt; 2' deoxythymidine-5'-(3'-phosphorothioate

&lt;400&gt; 11

acctatcctg gacctatcgt att

23

&lt;210&gt; 12

&lt;211&gt; 22

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 12

gcccaaaaag acagacagaa ag

22

&lt;210&gt; 13

&lt;211&gt; 23

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 13

agaaagacag acagggtggat aga

23

<210> 14

<211> 24

<212> DNA

<213> Homo sapiens

<220>

<221> misc feature

<222> (1)

<223> Biotinylated

<220>

<221> misc feature

<222> (23)

<223> 3'-deoxythymidine 5'-(S)-phosphorothioate

<400> 14

1ctctgttct gtcttcttat gata

24

<210> 15

<211> 21

<212> DNA

<213> Homo sapiens

<400> 15

tgagtgcaca atggagctt t

21

<210> 15

<211> 25

<212> DNA

<213> Homo sapiens

<220>

<221> misc feature

<222> (1)

<223> Biotinylated

<220>

<221> misc\_feature

<222> (25)

<223> 2'-deoxythymidine-5'-(S)-phosphorochanoate

<400> 16

gcttcacaat aacagttgct actall

26

<210> 17

<211> 23

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<222> (1)

<223> Niotinylated

<220>

<221> misc\_feature

<222> (13)

<223> 2' deoxythymidine-5'-(S)-phosphorotrioate

<400> 17

cccaagtgaat ttgccttcta

20

<210> 18

<211> 29

<212> DNA

<213> Homo sapiens

<400> 18

gtagatagac tggatagata gacgataga

29

<210> 19

<211> 24

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<222> (1)

<223> Biotinylated

<220>

<221> misc\_feature

<222> (23)

<223> 2'-deoxythymidine-5'-(S)-phosphorothioate

<400> 19

gtggtttaga tagatagata ggtu

24

<210> 20

<211> 20

<212> DNA

<213> Homo sapiens

<400> 20

ggttaaggag agtgcacta

26

<210> 21

<211> 19

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<222> (1)

<223> Biotinylated

<220>

<221> misc feature

<222> (15)

<223> 2'-deoxythymidine-5'-(S)-phosphorothioate

&lt;400&gt; 21

cagagcaggc ctccatctg

19

&lt;210&gt; 22

&lt;211&gt; 18

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 22

ccatcttagt gcctgttc

18

&lt;210&gt; 23

&lt;211&gt; 23

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (1)

&lt;223&gt; N-acylated

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (12)

&lt;223&gt; 2'-deoxythymidine-5'-(S)-phosphorothioate

&lt;400&gt; 23

ttaggagaca agcatagcag ttc

23

&lt;210&gt; 24

&lt;211&gt; 22

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 24

gcgaaagaat gagacacat ct

22

<210> 25

<211> 27

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<222> (1)

<223> Biotinylated

<220>

<221> misc\_feature

<222> (26)

<223> 2'-deoxythymidine-5'-(S) phosphorothioate

<400> 25

aaaattaggc atatttaca gctcgtt

27

<210> 26

<211> 25

<212> DNA

<213> Homo sapiens

<400> 26

tctgtaatcg ccagcaaaa agacu

25

<210> 27

<211> 29

<212> DNA

<213> Homo sapiens

<220>

<221> misc feature

<222> (1)

<223> Biotinylated

<220>

<221> misc\_feature

<222> (28)

<223> 2'-deoxythymidine-5'-(S) phosphorothioate

<400> 27

gtctccatcc ttgtctcctat ctctatctctg

28

<210> 28

<211> 22

<212> DNA

<213> Homo sapiens

<400> 28

gagaaggggc tgaattctgtc tt

22

<210> 29

<211> 20

<212> DNA

<213> Homo sapiens

<400> 29

ctgtgtctctc cclatttacc

20

<210> 30

<211> 21

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<222> (1)

<223> Biotinylated

<220>

<221> misc feature

<222> (20)

<223> 2'-deoxythymidine-5'-(S)-phosphorothioate

<400> 30

gggaacacag actccatggt g

21

<210> 31

<211> 22

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<222> (1)

<223> Biotinylated

<220>

<221> misc\_feature

<222> (2)

<223> 2'-deoxythymidine-5'-(S)-phosphorothioate

<400> 31

cttaggggaa ctcactgaa tg

22

<210> 32

<211> 22

<212> DNA

<213> Homo sapiens

<400> 32

gtccttgaca ggcTLLALL: gc

22

<210> 33

<211> 22

<212> DNA

<213> Homo sapiens

<400> 33



gtgtcagacc ctgttctaaag ta

22

<210> 34

<211> 20

<212> DNA

<213> Homo sapiens

<400> 34

atgaaatcaa cagaggcttg

23

<210> 35

<211> 25

<212> DNA

<213> Homo sapiens

<400> 35

ctctctggtt tctttatgta atlctt

24

<210> 36

<211> 24

<212> DNA

<213> Homo sapiens

<400> 36

tgtcatagtt tagaagaaac taac

25

<210> 37

<211> 27

<212> DNA

<213> Homo sapiens

<400> 37

tttgtatllc agtgtacat tcttacc

26

<210> 38

<211> 23

<212> DNA

<213> Homo sapiens

<400> 38

cccatcctaac gccatctctgt att

23

<210> 39

<211> 24

<212> DNA

<213> Homo sapiens

<400> 39

tcctctgtttt gtctctcttctt gata

24

<210> 40

<211> 26

<212> DNA

<213> Homo sapiens

<400> 40

gtctttacatt aacagcttctt actart

26

<210> 41

<211> 20

<212> DNA

<213> Homo sapiens

<400> 41

cccaagtga tlyccttctta

20

<210> 42

<211> 24

<212> DNA

<213> Homo sapiens

<400> 42

gtgtttttaga tagatagata ggtta

24

<210> 43

<211> 13

<212> DNA

<213> Homo sapiens

<400> 43

cagagcaaga cttcatctg

19

<210> 44

<211> 23

<212> DNA

<213> Homo sapiens

<400> 44

ctaggagaca aggatagcag ttc

23

<210> 45

<211> 27

<212> DNA

<213> Homo sapiens

<400> 45

aaacttaggc atattacaa gctagtc

27

<210> 46

<211> 29

<212> DNA

<213> Homo sapiens

<400> 46

gtctccatct ttgtctctat cttcatctg

29

<210> 47

<211> 21

<212> DNA

<213> Homo sapiens

&lt;400&gt; 47

gggaacacag actccatggt g

21

&lt;210&gt; 48

&lt;211&gt; 22

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 48

ctttaggggac cctccactgaa ag

22

&lt;210&gt; 49

&lt;211&gt; 21

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 49

gltggttaga cctttagcga t

21

&lt;210&gt; 50

&lt;211&gt; 16

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 50

gagtgagagt ccgtctca

18

&lt;210&gt; 51

&lt;211&gt; 24

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 51

ttccatcata caccacatc tgtc

24

&lt;210&gt; 52

&lt;211&gt; 22

<212> DNA

<213> Homo sapiens

<400> 52

ataquaggyat aggtaggcag gc

22

<210> 53

<211> 19

<212> DNA

<213> Homo sapiens

<400> 53

ttttttcttgL atcaacatt

19

<210> 54

<211> 22

<212> DNA

<213> Homo sapiens

<400> 54

aaactacca atcccatcc tt

22

<210> 55

<211> 23

<212> DNA

<213> Homo sapiens

<400> 55

tggtcttcta cttgtgtcaa tac

23

<210> 56

<211> 23

<212> DNA

<213> Homo sapiens

<400> 56

tgtctcatag aacagacata cat

23

<210> 57

<211> 23

<212> DNA

<213> Homo sapiens

<400> 57

ctggatctct tggctatagt aaa

23

<210> 58

<211> 23

<212> DNA

<213> Homo sapiens

<400> 58

aagctggtag agagatacaa aga

23

<210> 59

<211> 18

<212> DNA

<213> Homo sapiens

<400> 59

agccacaccc ataacctt

18

<210> 60

<211> 22

<212> DNA

<213> Homo sapiens

<400> 60

gaatgcagag aaagcymtc ta

22

<210> 61

<211> 8

<212> DNA

<213> Homo sapiens

&lt;400&gt; 61

agaaatggct tggccttg

19

&lt;210&gt; 62

&lt;211&gt; 18

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 62

taaaggattg caggaggg

18

&lt;210&gt; 63

&lt;211&gt; 23

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 63

gaataagatt ctgttgaggg aaa

23

&lt;210&gt; 64

&lt;211&gt; 24

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 64

aatcttctct cttctacct ctct

24

&lt;210&gt; 65

&lt;211&gt; 26

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 65

aggyctgatg aggtgaaata

26

&lt;210&gt; 66

<211> 24

<212> DNA

<213> Homo sapiens

<400> 66

tcacacacaa cacatataag atga

24

<210> 67

<211> 25

<212> DNA

<213> Homo sapiens

<400> 67

catacatttg tagatcgata gaaga

25

<210> 68

<211> 23

<212> DNA

<213> Homo sapiens

<400> 68

gggtctccca gagaacaga atc

23

<210> 69

<211> 29

<212> DNA

<213> Homo sapiens

<400> 69

cagacagat agatagatat gtacataaa

29

<210> 70

<211> 24

<212> DNA

<213> Homo sapiens

<400> 70



tcaaagagtg acaagaaac taa

24

<210> 71

<211> 24

<212> DNA

<213> Homo sapiens

<400> 71

ccatccatct atcatctatt tttt

24

<210> 72

<211> 23

<212> DNA

<213> Homo sapiens

<400> 72

acctacatta gtctgtgllc ttt

23

<210> 73

<211> 21

<212> DNA

<213> Homo sapiens

<400> 73

aagaagaat gaccttggg attt

24

<210> 74

<211> 19

<212> DNA

<213> Homo sapiens

<400> 74

gggcgacaga gcaagactc

19

<210> 75

<211> 20

<212> DNA

<213> Homo sapiens

<400> 75

ctggagtcgca agctgaacta

20

<210> 76

<211> 23

<212> DNA

<213> Homo sapiens

<400> 76

ctgagtgaca gaggtagaac cng

21

<210> 77

<211> 19

<212> DNA

<213> Homo sapiens

<400> 77

atcaatggt gcataggca

19

<210> 78

<211> 23

<212> DNA

<213> Homo sapiens

<400> 78

gcttcatat cacttgagct aat

23

<210> 79

<211> 25

<212> DNA

<213> Homo sapiens

<400> 79

gcttactga ctactacal aacga

25

<210> 80

<211> 21

<212> DNA

<213> Homo sapiens

<400> 80

gagcaagact gaattctttaa a

21

<210> 81

<211> 23

<212> DNA

<213> Homo sapiens

<400> 81

Lygaaaaata ctctgggaag ata

23

<210> 82

<211> 22

<212> DNA

<213> Homo sapiens

<400> 82

cctgllgaca ctctcaaaac ca

22

<210> 83

<211> 16

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<222> (1)

<223> Biotinylated

<220>

<221> misc\_feature

<222> (16)

<223> 2'-deoxythymidine-5'-(S)-phosphorothioate

<400> 83

gagcgaagat ccgtctca

18

<210> 84

<211> 24

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<222> (1)

<223> Biotinylated

<220>

<221> misc\_feature

<222> (23)

<223> 2'-deoxythymidine-5'-(S)-phosphorothioate

<400> 84

tccatcaca caccatata tgta

24

<210> 85

<211> 19

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<222> (1)

<223> Biotinylated

<220>

<221> misc\_feature

<222> (18)

<223> 2'-deoxythymidine-5'-(S)-phosphorothioate

&lt;400&gt; 85

ttttttcttgt atcaaat

19

&lt;210&gt; 86

&lt;211&gt; 23

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (1)

&lt;223&gt; Biotinylated

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (21)

&lt;223&gt; 2'-deoxythymidine-5'-(3')-phosphorothioate

&lt;400&gt; 86

tggctctctca ctctgtctcaa tac

23

&lt;210&gt; 87

&lt;211&gt; 23

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc feature

&lt;222&gt; (1)

&lt;223&gt; Biotinylated

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (20)

&lt;223&gt; 2'-deoxythymidine-5'-(3') phosphorothioate

<400> 87

ctggacttct tggllulagc aag

21

<210> 88

<211> 22

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<222> (1)

<223> Biotinylated

<220>

<221> misc\_feature

<222> (21)

<223> 2'-deoxythymidine-5'-(S)-phosphorothioate

<400> 88

gaatgcagcg aaggagatc ta

22

<210> 89

<211> 18

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<222> (1)

<223> Biotinylated

<220>

<221> misc\_feature

<222> (17)

<223> 2'-deoxythymidine-5'-(S)-phosphorothioate

<400> 89

agaaatggct tggccttg

18

&lt;210&gt; 50

&lt;211&gt; 24

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (1)

&lt;223&gt; Biotinylated

&lt;220&gt;

&lt;221&gt; misc feature

&lt;222&gt; (22)

&lt;223&gt; 2'-deoxythymidine-5'-(S)-phosphorothioate

&lt;400&gt; 90

aatctctctct attctacct ctc

24

&lt;210&gt; 91

&lt;211&gt; 20

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc feature

&lt;222&gt; (1)

&lt;223&gt; Biotinylated

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (19)

&lt;223&gt; 2'-deoxythymidine-5'-(S)-phosphorothioate

&lt;400&gt; 91

agggtgaly aggtgaata

26

<210> 32

<211> 23

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<222> (1)

<223> Biotinylated

<220>

<221> misc\_feature

<222> (22)

<223> 2'-deoxythymidine-5'-(S)-phosphorothioate

<400> 32

gagLLGcna gagaacaga atc

23

<210> 33

<211> 29

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<222> (1)

<223> Biotinylated

<220>

<221> misc\_feature

<222> (26)

<223> 2' deoxythymidine-5'-(S)-phosphorothioate

<400> 33

cagacaaat agatagaLac gtacataca

29



<210> 94

<211> 24

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<222> (1)

<223> Biotinylated

<220>

<221> misc\_feature

<222> (23)

<223> 2'-deoxythymidine-5'-(3)-phosphorothioate

<400> 94

ccatccatcc atcctccatt tact

24

<210> 95

<211> 24

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<222> (1)

<223> Biotinylated

<220>

<221> misc\_feature

<222> (23)

<223> 2'-deoxythymidine-5'-(3)-phosphorothioate

<400> 95

aagaaagaat gaccc.Lyga atttt

24

<210> 96

<211> 20

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<222> (1)

<223> Biotinylated

<220>

<221> misc\_feature

<222> (19)

<223> 2'-deoxythymidine-5'-(3')-phosphorothioate

<400> 96

tggagtcgca agc-gaacta

20

<210> 97

<211> 19

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<222> (1)

<223> Biotinylated

<220>

<221> misc\_feature

<222> (18)

<223> 2'-deoxythymidine-5'-(3')-phosphorothioate

<400> 97

atcaatggat gcataggta

19

<210> 98

<211> 25

<212> DNA

<213> Homo sapiens

<220>

<221> misc feature

<222> (1)

<223> Biotinylated

<220>

<221> misc feature

<222> (20)

<223> 2'-deoxythymidine-5'-(3')-phosphorothioate

<400> 58

ggcctactga ctactacat aacga

25

<210> 99

<211> 23

<212> DNA

<213> Homo sapiens

<220>

<221> misc feature

<222> (1)

<223> Biotinylated

<220>

<221> misc feature

<222> (22)

<223> 2'-deoxythymidine-5'-(3')-phosphorothioate

<400> 99

tgguaaaata ttctgggaag ata

23

<210> 100

<211> 23

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<222> (1)

<223> Biotinylated

<220>

<221> misc feature

<222> (21)

<223> 2'-deoxythymidine-5'-(S)-phosphorothioate

<400> 100

acagtaactg cctccataga tag

21

<210> 101

<211> 22

<212> DNA

<213> Homo sapiens

<400> 101

gcctgttcct cccctatttc cc

22

<210> 102

<211> 23

<212> DNA

<213> Homo sapiens

<400> 102

aggtcacagg gaacacagac tcc

23

<210> 103

<211> 23

<212> DNA

<213> Homo sapiens

<220>

<221> misc. feature

<222> (1)

<223> Nicotylated

<220>

<221> misc. feature

<222> (21)

<223> 2'-deoxythymidine-5'-(S) phosphorothioate

<400> 103

aggtcacagg gaacacagac acc

23

<210> 104

<211> 315

<212> DNA

<213> Homo sapiens

<400> 104

aacctgagtc tgcacaggac tagcaggctt ctacacaccc tctgtctcag tcttcctacc 60  
 tgaataatga agatallat agtaactgac ttcataagata gaagatagat agatcagata 120  
 gatagataga tagatagata gatagataga tagatagata gataggagat acttagaata 180  
 gggctctgac caggaaatgc tctccaaagt tgcaccaggga gatagatctt gagaaagctc 240  
 agctcagatc catgagggtt gggtaggaac ctggagagct gagaaaggc tgaagatagg 300  
 cagtggctgt tggaa 315

<210> 105

<211> 307

<212> DNA

<213> Homo sapiens

<400> 105

tctaattaaa gtgggtgcc agalautctg tactaataaa agtalaatt aatagcaagt 60  
 ukgtganaag ggtgattttt ctcttttgta tcttatgta atattttgaa gatagataga 120  
 tagatagata gatagataga tagatagata gatagataga tagaggtala aataggata 180  
 cagatutagn caaaatgtt gttaactgtg gctatgattg gaataacttg gctaaaaagc 240  
 gaaaagcct tctctgnga yaggcaatta ctcttttctt taggaactnc ctoancagtc 300  
 tactngc 307

<210> 136

<211> 334

<212> DNA

<213> Homo sapiens

<400> 136

```

aatttttggc attttttttag agacgggggll Laaccatggt ggtagcgcacg actatggagt 50
taatttaagg taattatata taaagggLat gatagaacac ttgtcatagt ttayaaagaa 120
ctaacgatay ulugatagat agataatag atagalaayal agatagatag atagacagat 180
tgatagttll tttttatctc ac_aaatagt ctatagtaaa catttanta ccaatattlg 240
gtgcaattct gtcaatgagz ulaaatgtgg aatcgttata atllttaaga atatatattc 300
uu_cagagat ttgatacet cagattttaa ygc 334

```

<210> 137

<211> 340

<212> DNA

<213> Homo sapiens

<400> 137

```

tggcaattta tagtatattt tggattccat ggtacatcc gtactatctt atctatctat 50
ctatctatct atctatctat ctatctatcl attcccccac ttgaaataa totacagga 120
aggtaaataa attauggcat attcacgcac tgggatacgn Lacagtgcg aadctgaact 180
aattatagct aggtgaaact atactcatgt acacaattll; gtaaaagaaa cluggaaca 240
gaatacatat gytttttgnc agctgLyca tttaaatcc ccaacaaac Lyucaggg 300
ttcagnttcl ccacatnctt gtcaacatzz tglLaltttg 340

```

<210> 138

<211> 286

<212> DNA

<213> Homo sapiens

<400> 138

```

tgggatgggt agctggacat ggtatcaozz aagtctggga tgtggaggag agllcatncc 50
tttagtgggc alcuagtact ctctggactc tgaocentct aagocetacc tgratttacc 120
aatacattat ctatctatct alclalttat ctatctatct atctatctat ctatctalca 180
atcatctatc tatcttllLy tctgtctctt tgggtgcoct alygctcaac ccaggttga 240

```

gg ggagatt cgacccacuu ttcaagctct ctgaatatgc ttlgaa 266

<210> 109

<211> 426

<212> DNA

<213> Homo sapiens

<400> 109

atgggtgccc ccacggctgc acggggagga tgactgtntt cccuclctca gtcnrgccga 60  
 ggtgcccgac agccccggac ccaggagctg ggggggtctaa gaccttctaa aaagtgtacn 120  
 agtgcacgcl gctcgtctgt cacaacctca aatgcagaaa agcactgaaa gacgaalcca 180  
 gaaacaccca gtccaccatt ctatatggga gcauucuaag gcagatocca agctctccct 240  
 ctcccccaga ccaatcacga cagucugaca ggttggataga tagatagala gctagataga 300  
 tagatagaaa garagatctt attgaaagac aaaaacagaga tggutgataq aacacatgct 360  
 acagatgcac acacaaacgt aatatggcat aaaaatggga tcaactcttg targgtcgtt 420  
 nttacc 426

<210> 110

<211> 350

<212> DNA

<213> Homo sapiens

<400> 110

aggttaaggc tgcagtgcgc calgntcatg ccactgcact ccactccgag tgacaaatly 50  
 agacctctgt tcagaaagaa agaaagaaag aaagaaagaa agaaagaaag aatgaaagaa 120  
 agaaagtaag aaaaagagag ggaagaaag agaaagaaag aatgaaagaa agaaagaaag 180  
 attgtaaagc atctccacac aacagagaaag ttaatttcaa ttttaacatg taaagaaag 240  
 agagaaagca acatgtccac cttaggctga cgttttgttt atttgtgttg tggctggtag 300  
 tgggtgtgtt tatttttcaa gtagnttctc caatactcca ttaacatctt 350

<210> 111

<211> 328

<212> DNA

<213> Homo sapiens

<400> 111

ctaccatctc tcttggaaag caaagacaga gcaagacatc tcacatggct agagcagcag 60

```

caagagaag ataggggagc tcttagatgg tctgttatgg gacttttctc agtctccala 120
aatatgtgag tcaatccccc aagtgaaatg cttctatctt alclalatat ctgtctgtct 180
gtctgtctgt ctgtctatct acctacatcl acctatctat catctatcta tctatctate 240
tatctatctc tctatctate tctgtctat. ctatccagtc tatccacctc clallagact 300
gtctctggag aacatcgact aatacaacat ctllaatata tccaggttta atttcaagtl 360
atatcatacc acctccatca clatataaaa cctttagctg tttctccctt ctccagly-ct 420
atggclayta allllctart ggggtgcaga cactaatlll tttctgcta agtggtgaat 480
attttttata tctttaaaaa tttttlyag tgttgaltcg ggtaaagt 528

```

&lt;210&gt; 112

&lt;211&gt; 194

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 112

```

ctactcagtt tctgttatag tgltttttaa tatatatata gtattatata tataglylta 50
tatatatata tgltttttaa ttagtagata gtagataga tagatagata gtagataga 120
tagatagata gtagataga tatagtagaa tttctcttaa cttagataga ctcttctac 180
tactacatg cat 194

```

&lt;210&gt; 113

&lt;211&gt; 320

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 113

```

cacttgaacc cgggaggtgg aggttguaat ccagctcttg caacagagca aucttcato 50
tgaaugatag aaagatgaa gaaagaaaga aagaaagaaa gaaagagtaa aaqaaaaaa 120
ttaaattctt agggggaaaa ttttcaatt tttagaacatg cactaaaaatg atttccagag 180
aaaaccaagc glllctctt aatctgcatg gcattatlaa agatgtctac tctctctct 240
tggggctagg catcccttc ctccaggaag tctgttgtt aggggggggc tctggctctg 300
cgatgattca ggaatgcaga 320

```

&lt;210&gt; 114

&lt;211&gt; 330

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens



&lt;400&gt; 114

```

gggatctccc tatggattgg aagggggcg tgaatagag gagtcagggg tcaactctgg 60
gatttggcct ggagcagctg gaagatggag tggctgttaa ttcattgtag gagggtctg 120
ggaagaagag tttaggaga caaggalagc agtcaccta tttatttatt ttttatcta 180
tttatcatt tatctattta gagatgagt ctatctctc cgcaggctg gactgagtg 240
gagcagctt ggctcactgc aacctcacc tccaggctc aagcagctt ttggctcag 300
ctcccgagt agccagtag ctgggaactc

```

330

&lt;210&gt; 115

&lt;211&gt; 192

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 115

```

gccccatag tttgaactc acagatcaa ctgtaacca aataaaalia ggcattctta 60
caagcagll tcttctctc ttttctctc ttcttctct ctctctctct tcttctctt 120
tttctctct tcttctctc ctcttctct ctcttctct ttttctctt gctggcaall 180
acagacaaat ca

```

192

&lt;210&gt; 116

&lt;211&gt; 320

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 116

```

aggatactt ttctctcag aatagttaga ttaggtata ccactctgat gtgacacta 60
glttaactag aacttatctt cgttaaatc gtctctatc caatctctgt ctccatctt 120
gtctctatct ctatctgtct acctctctc atctatctat ctatctctc tctatctat 180
ctatctctc atctaaagca aatcctgct ctctctctt tctctctct gagaccatg 240
acaggggtga gacaaagaa tggcaggaa tgggagtg gattatctgt ggcalaagg 300
aactttacag aactaggtc

```

320

&lt;210&gt; 117

&lt;211&gt; 300

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 117

ggccttccca ggctctagaa gcagcucag gtggggggga ctgggcaaat agggggcaaa 60  
 attcaaaagg tatctggggt clygggtgat tcccattggc ctgttcccluc attttctccc 120  
 lcullcatnn attcattcat tcaatcatcc attcaatcac catggagatct gtgttccctg 180  
 tgaactgcac toggaaagac tgtgtacagc ggactgtgtg ggccaggctg gataacaggg 240  
 agcttttcag ccacacaggag gggctcttcg tgcclealtg ggcacccaga accttgggct 300

&lt;210&gt; 118

&lt;211&gt; 300

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 118

agcaccaga accgtcgact qgcacagaa aggcacttgg ggaacacctca ctgaatgaat 60  
 gaattgaatga atgaalgaat gaatgaatga atgaatgttt ggscacataa acyulgaana 120  
 ggaacagaagg gccluycggg aagggaacac gactaayac agcgacacag ccgaactngtg 180  
 ltcagaaagac ctgggatttg acctgagagc ttcacatttg gatgaatctt ttaattaac 240  
 tttgtgggtc ccagtctctc ccclygucgc ccaggacagt agagtcaaac lcaagtctga 300

&lt;210&gt; 119

&lt;211&gt; 143

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 119

gtgagttaga cgttcagcga tatucacata tlatgaaca ttatlaltat tattattatt 60  
 attatlaltat ttattattat tattattatt lyagacggac tctcgtctctg tgcgccagge 120  
 tggagcgcag tggctgcga= ctg 143

&lt;210&gt; 120

&lt;211&gt; 279

&lt;212&gt; RNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 120

ccattcalle uatcacacac caatatctgt clylctgtct acctatctat ctatctatct 60

```

atctatctat ctatctgct atctgcttgc ctacctatcc ctctatggca attgcttgc 120
accaggagga ttttattccc agggagatct tggctatgng tgacaacaal llttttgggt 180
gtcaccaatg ggatgaatgt tactggcatc Lyy..gggtgg agcccagaga tggctgctcaa 240
caccctacag tgcacaagac agcccacuu caagaatc 279

```

&lt;210&gt; 121

&lt;211&gt; 263

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 121

```

ccattaacct agclllluuu macaactaal ttgatttcaa glytttgcct ttcaaaagcc 60
aagaaggaaa acaatttttt ttctctgta_c accatttail taktattatt attattatca 120
ctattatcat tactattatt attalllakt actaagggaat gggattggta gylltaattga 180
ccccclgllt ttgacttctt tgagatattt ccagactact ttccactctg aacttaggaa 240
tttactctgc atcaactggg ccl 263

```

&lt;210&gt; 122

&lt;211&gt; 131

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 122

```

gtggtctctt acctgtgtca atacagatag atagatagat agatagatag atagatagat 60
agatagatag atagatagat agulutghat gtcttttcta tgagacalac ctcatcttcl 120
ggulttgagt t 131

```

&lt;210&gt; 123

&lt;211&gt; 372

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 123

```

catgngcccc caaagcgnag caaactluac ccagtgtcac aaaatggccl lluaingnatt 60
actccclouut tgnccaccca cctatacttc aclytctgga ttctctggth atagtaaatc 120
tagatctatc tatctatola ctatctatc tatctatcta tctatctatc tatctgltg 180
tctctctacc agcclltttha aactgtctct aattgttcaa ttatctatata atgagaaat 240

```

ggctatatatt tectqagngc rgnnttaucca tagtagtggc aabqagttgc agnancaggg 360  
 ocaacattgt caactcttgg ttatluccnc atgttttnc utttncnta aatttcaatt 360  
 ctaggnggtc ca 372

<210> 124

<211> 242

<212> DNA

<213> Homo sapiens

<400> 124

agctacagca aactccatgt gacaaagcc acacccatna ctttttncct ctgatataga 60  
 agatagatga tagatagala gataataga tagatagata gatagataga laggatagata 120  
 garatagatt ctctttctct gcatttccat ctatattctct gtctttctct taathatggg 180  
 taactcttag cctggcaggc taccatggca agacaaactt lattctctct ttctctctggc 240

<210> 125

<211> 325

<212> DNA

<213> Homo sapiens

<400> 125

gtgggaggaa gccagtggat ttggaaacag aaatgggttg gcttttctg cctgcctgcc 60  
 tgcctgcttl ucllccctnc tttcttctct cttctcttcc ttccllcll cctctctgca 120  
 atcctttaac ttactgaata actcattatt atgggtcncd lgcaggtanc atgtatggta 180  
 ctagggatgt aggcattgaac actgaatggg gctcttggga cctgcattct ggtaggaaaa 240  
 ggggtlgagc agggatagaag ccaatcaatg tctcaatgg aaacagtctt aagtgtctagc 300  
 aaagaaatgaa cgtattgatg tcaca 325

<210> 126

<211> 269

<212> DNA

<213> Homo sapiens

<400> 126

aaagctaluu lgtatccact gactccagc ctgggcaaca gaataagatt ctgttggagg 60  
 aaatgaagta ggaagggaag aaggaaaggaa ggaagggaagg aaggaaaggaa ggaagggaag 120  
 aagtagaag agagaagaa LLlLuttcggg taaagggtgc accaaaatct cagaaalcac 180

tgctaaagaa cttatttcattg taaccaatcac cactctgtccc taaaaacct attgaaalaa 240  
 anacagccag mmyuyugaa agaggmnga 259

<210> 127

<211> 377

<212> DNA

<213> Homo sapiens

<400> 127

aaagtcttca aagatccctg aagttggttca taagccagca tttttttttt tctaaaggagg 60  
 caacaaaaga tttttttttt gtacagctaa ttgtgactct gaacccagag tttttttttt 120  
 gctttttttt aacccagagat ttattttgca ttalagaagg ttctctagga tttttttttt 180  
 tttttttttt ggagatagtt atattttttt ctgttttttt gctgatgagg tgaatatatt 240  
 gcaaaacaaat ctatcttttt ttattttttt ctatcttttt atcttttttt ctatcttttt 300  
 tttttttttt ctatcttttt atattttttt ttgttttttt ttgttttttt atcttttttt 360  
 tttttttttt atatttttt 377

<210> 128

<211> 344

<212> DNA

<213> Homo sapiens

<400> 128

tttttttttt aacttttttt gtgttttttt ttgttttttt gacttttttt ttcttttttt 60  
 aacttttttt ttgttttttt ttgttttttt gctttttttt ttgttttttt ttgttttttt 120  
 gacttttttt ttgttttttt ttgttttttt attttttttt ttgttttttt ttgttttttt 180  
 attttttttt ttgttttttt ttgttttttt ttgttttttt ttgttttttt ttgttttttt 240  
 ttgttttttt ttgttttttt ttgttttttt ttgttttttt ttgttttttt ttgttttttt 300  
 ttgttttttt ttgttttttt ttgttttttt ttgttttttt ttgttttttt ttgttttttt 344

<210> 129

<211> 372

<212> DNA

<213> Homo sapiens

<400> 129

aaagcttttt ttgttttttt ttgttttttt ttgttttttt ttgttttttt ttgttttttt 60

ctagatagat ggacagacta gatagataga taagtacata cataagatag atagatagar 120  
 agulugulug atagatagat agatagatag atagatagat agagacagat ttaaatatatt 160  
 tgggacatct tagttctctt gtaactctct gaattgggac tataaaaaat actctctttac 240  
 tatcacaaga ggulugugga ccaatataat tgcactgct ggtctctaaa agtgaacgcc 300  
 aggtacaagg gttaccatta ctccctcttg ggctcaggat ctgctctgct tgcagccacc 360  
 actcaactgta ct 372

<210> 130

<211> 355

<212> DNA

<213> Homo sapiens

<400> 130

ttacctaat tggctctaga ccatacctaa atctctctct ctctctctct 60  
 ctctccctct ttacaggga agtctctct agctatctac ccaatttgag ttgagcttt 120  
 ttgagagaaa gaattctct atgtctctct atctatccat ccatacctct atctctctct 180  
 cctctctct ctctctctct atctctctct ctctctctct atctctctct ctctctctct 240  
 gaacacaga taattgaggt gatactagg atctctctct cactaagcat agtctctct 300  
 cctgaacccc agaggaggaa ctctctctct ctctctctct aggggctct 355

<210> 131

<211> 320

<212> DNA

<213> Homo sapiens

<400> 131

taattgtaat atttctctaa agtctctct ccaattggt tctgaggtgg tctctctct 60  
 tctctgagaa tctctctct tctctctct gagacacaa taagaaaga atgacccctg 120  
 gaattctct taattctct cctctctct tctctctct tctctctct tctctctct 180  
 gagctctct ctctctctct cctctctct caatggcag atctctctct actgcaattt 240  
 tctctctct ggctctctct atctctctct ctctctctct caagtagctg ggatctctct 300  
 cgtctctct ctctctctct 320

<210> 132

<211> 159

<212> DNA

<213> Homo sapiens

&lt;400&gt; 132

ctggcglcgc uagcagaact agcgtttctt tttctllctt tttctttct tttctttct 60  
 tttctttctt tttctttctt tcaagacagg llclacnctt gtcacacagg cttagagtgc 120  
 gtggcgcaat caaggllctt tgnagcctca acttctctg 159

&lt;210&gt; 133

&lt;211&gt; 229

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 133

uucuggatca atggatgcac aggttagatg atagatagat agatagatag atagatagat 60  
 agatagatag atagatagac agacagacag aagacagac agatgagagg ggattttatta 120  
 gaggatttag ctcauglyat atggaggctt aaaaatctca tgcagtcctt cctgcaagct 180  
 ggagaccag ggaactaggt agcatggctc agtccaggct taaaagcca 229

&lt;210&gt; 134

&lt;211&gt; 379

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 134

agcacaatat aacttcacag attgaacaca tccatgtaac agcacaacag attaagaaac 60  
 agagcatgac tggacaatc tcatgcllct ttttagacac taccgttgac ccttaantaa 120  
 tttggggull uggggtgcag tgganaatcc aagtalaatt ttgtctcctt gaaantgtta 180  
 ctagttaatag cctactgttg actggaagcc llactgactt actacataac tacacacaca 240  
 cacacacaca caccacacac cacacacaca cacacacata talakatttt gagatgcagt 300  
 ctctgctctgt tggccaggct ggagtncaat ggcatgalec aggttcaatg caacctccgc 360  
 cttctgggtt caagcggct 379

&lt;210&gt; 135

&lt;211&gt; 387

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 135

gaattataac cgttaattgat tcatagcagc aattgcaaaa ttctattttg tggaaaaaa 60  
Llclgygaag atakttaaca tgnacacac acacacacac acacacacac acacacacac 120  
gatgacatg gttttaaaaa tylaacagg ttcctttgct qgaggaattc ccagtgtttt 180  
tgttatagga atccllcaclg ggaataaagt gataatagca gtggtaatgg aaatgtttta 240  
ttgactgctt aaactgaagt canacaagca ttatctcaat tttttakaa aaattattta 300  
atcllcaaaa cagacctggc cagtaggtan aattatgtgg tacacagatg agaaactgag 360  
gcttacagag atgacgataa cccagct 387